

Probabilistic Library

Scientific background

Version: 25.3.1
Revision: ----default----

9 January 2026

Probabilistic Library, Scientific background

Published and printed by:

Deltares
Boussinesqweg 1
2629 HV Delft
P.O. 177
2600 MH Delft
The Netherlands

telephone: +31 88 335 82 73
e-mail: [Information](#)
www: [Deltares](#)

For sales contact:

telephone: +31 88 335 81 88
e-mail: [Sales](#)
www: [Sales & Support](#)

For support contact:

telephone: +31 88 335 81 00
e-mail: [Support](#)
www: [Sales & Support](#)

Copyright © 2026 Deltares

All rights reserved. No part of this document may be reproduced in any form by print, photo print, photo copy, microfilm or any other means, without written permission from the publisher: Deltares.

Contents

List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Overview of this document	1
2 Statistics	2
2.1 Distributions	2
2.1.1 Generic description of the application of distribution functions	2
2.1.2 Distribution properties	6
2.1.3 Distribution types	7
2.1.3.1 Deterministic distribution	7
2.1.3.2 Normal distribution	7
2.1.3.3 Standard normal distribution	9
2.1.3.4 Log normal distribution	10
2.1.3.5 Uniform distribution	11
2.1.3.6 Triangular distribution	12
2.1.3.7 Trapezoidal distribution	13
2.1.3.8 Exponential distribution	13
2.1.3.9 Gumbel distribution	14
2.1.3.10 Weibull distribution	15
2.1.3.11 Conditional Weibull distribution	15
2.1.3.12 Frechet distribution	16
2.1.3.13 Generalized Extreme Value distribution	16
2.1.3.14 Rayleigh distribution	17
2.1.3.15 Rayleigh- N distribution	17
2.1.3.16 Pareto distribution	17
2.1.3.17 Generalized Pareto distribution	18
2.1.3.18 Student's T distribution	18
2.1.3.19 Gamma distribution	19
2.1.3.20 Beta distribution	19
2.1.3.21 Poisson distribution	20
2.1.3.22 Discrete distribution	20
2.1.3.23 Bernoulli distribution	20
2.1.3.24 CDF curve distribution	21
2.1.3.25 Histogram distribution	21
2.1.3.26 Inverted distribution	22
2.1.3.27 Truncated distribution	22
2.1.4 Prior distribution	22
2.1.5 Design values	23
2.1.6 Goodness of fit	23
2.2 Correlations	24
2.2.1 Correlation factor	24

2.2.1.1	Distance based correlation factor	24
2.2.2	Processing of correlation factors	25
2.2.3	Archimedean Copulas	25
2.3	Conversion between standard normal variables (U -domain) and real variables (X -domain)	26
3	Sensitivity	29
3.1	Single value variation	29
3.2	Sobol indices	29
4	Output uncertainty	31
4.1	Numerical Integration	31
4.2	Crude Monte Carlo	31
4.3	Importance Sampling	31
4.4	Directional Sampling	31
4.5	FORM	33
4.6	FOSM	34
5	Reliability	35
5.1	Probability of failure of a single component	35
5.2	Reliability index β	36
5.3	Reliability methods	39
5.3.1	Numerical Integration	39
5.3.2	Numerical Bisection	40
5.3.3	Crude Monte Carlo	40
5.3.3.1	Algorithm	40
5.3.3.2	Convergence	41
5.3.4	Importance Sampling	42
5.3.4.1	Algorithm	42
5.3.4.2	Convergence	43
5.3.4.3	Mean realization	44
5.3.5	Adaptive Importance Sampling	45
5.3.5.1	Algorithm	45
5.3.5.2	Auto break loops	46
5.3.6	Subset simulation	47
5.3.7	Directional Sampling	49
5.3.7.1	Algorithm	49
5.3.7.2	Convergence	49
5.3.8	Latin Hypercube	50
5.3.9	Cobyla	50
5.3.10	FORM	51
5.3.10.1	Algorithm	51
5.3.10.2	Convergence	53
5.3.10.3	Starting point	54
5.3.10.4	Loops	54
5.3.11	FORM then Directional Sampling	55
5.3.12	Directional Sampling then FORM	55
5.4	Contribution per variable	55
5.4.1	Center of gravity	55
5.4.2	Center of angles	56
5.4.3	Nearest to mean	56
5.5	System analysis	56
5.5.1	Equivalent planes	56
5.5.2	Combine methods	57

5.5.2.1	Directional sampling	57
5.5.2.2	Importance sampling	57
5.5.2.3	Hohenbichler	58
5.6	Upscaling for systems with identical components	59
5.6.1	Computational procedure	59
5.6.1.1	Detailed explanation	59
5.6.1.2	Equivalent α -values	62
5.6.2	Techniques for time and space dependent processes	66
5.6.2.1	Poisson counting process	67
5.6.2.2	Outcrossing	68
5.6.3	Spatial upscaling - from cross section to flood defence segment	70
5.6.3.1	Computing the failure probability	70
5.6.3.2	Computing equivalent α -values	73
A	Probabilistic computation techniques for system reliability	75
A.1	Linearisation of Z -functions	75
A.2	Failure probability for a single component	78
A.2.1	Introduction	78
A.2.2	Numerical Integration	79
A.2.3	Crude Monte Carlo	81
A.2.4	Monte Carlo Importance Sampling	83
A.2.4.1	General description	83
A.2.4.2	Implementation of the method of increased variance	85
A.2.5	Monte Carlo Directional Sampling	86
A.2.5.1	Determining the distance to the limit state	87
A.2.6	First-order reliability method (FORM)	89
A.2.7	Computing α values for other methods than FORM	97
A.2.8	Rationale	99
A.3	Combining failure probabilities for components - generic methods	100
A.3.1	Combining n system components: the Hohenbichler method	101
A.3.1.1	Introduction	101
A.3.1.2	Hohenbichler method for a system with fully correlated variables	103
A.3.1.3	Derivation of equivalent α -values	107
A.3.1.4	Hohenbichler method for the general case of partial correlation	109
A.3.1.5	Systems with an arbitrary number of components	113
B	Correlations between random variables	117
B.1	General description	117
B.2	Correlation models	118
B.2.1	Gaussian correlation model	119
B.2.2	Gaussian copula model	120
C	Conversions	122
C.1	Background	122
C.1.1	Floating point computations	122
C.1.2	The (inverse) cumulative distribution function (CDF)	122
C.1.3	The (inverse) of the standard normal distribution function	123
C.2	Conversion of reliability index to probability (and vice versa)	123
References		125

List of Tables

5.1	Values of reliability index β and associated probability of failure.	38
A.1	Computation techniques available in the Probabilistic Library for the computation of failure probability of a cross section of a longitudinal segment.	79
A.2	Required number of samples with crude Monte Carlo for combinations of the acceptable relative error ε and actual probability of failure P_f . The value of k in equation (A.18) is taken equal to 1.96.	83
A.3	Description of the Z -functions of Figure A.22.	109
A.4	Description of the Z -functions of Figure A.23	114

List of Figures

2.1	PDF (top), CDF (middle) and inverse CDF (bottom) of the standard normal distribution function.	3
2.2	Procedure for determining a load variable associated with a randomly selected standard normally distributed variable (u -value), for the case of uncorrelated variables.	4
2.3	Standard uniform distribution function (CDF).	5
2.4	Normal probability density function, with parameters μ and σ indicated	8
2.5	Illustration of the effect of parameters μ and σ	9
2.6	Standard normal distribution	10
2.7	Uniform probability density function, with parameters a and b indicated	11
2.8	Uniform cumulative distribution function, with parameters a and b indicated . .	12
2.9	Schematic view of the transformation between standard normal variable u_i and real world variable X by means of equal probability of (non-)exceedance.	27
2.10	Schematic view of correlated variables x_1 and x_2	28
4.1	Generation of first samples in random directions	32
4.2	The second/final step of iterations for the valid directions	33
5.1	Illustration of a river levee exposed to the threat of overflow.	36
5.2	Example of normally distributed S , R and Z functions with $\mu_R = 2$, $\mu_S = 6$, $\sigma_R = 1$, $\sigma_S = 1$, $\mu_Z = 4$ and $\sigma_Z = \sqrt{2}$	37
5.3	Schematic view of the relation between the reliability index β and the probability of failure, $P[Z < 0]$. Z is normally distributed with mean β and standard deviation 1	38
5.4	Numerical integration: realizations for two stochasts	39
5.5	Numerical Bisection: realizations for two random variables	40
5.6	Crude Monte Carlo: realizations for two random variables	41
5.7	Crude Monte Carlo convergence	42
5.8	Importance Sampling: realizations for two random variables	44
5.9	Adaptive Importance Sampling	45
5.10	Directional Sampling: realizations for two random variables	50
5.11	Schematic representation of FORM	51
5.12	FORM: example of iteration path to design point	54
5.13	Schematic view of a parallel system, series system and a combination of both. Components 1, 2 and 3 can be viewed as bridges. The systems fails if a passenger can not walk from left to right over the bridges.	56
5.14	Stochastic variation of the Z -function in time (left) and space (right)	67
5.15	Standard exponential distribution function: $F(T) = 1 - \exp(-T)$	68
5.16	Autocorrelation function, correlation within a dike section; in this picture, the correlation ρ is visualized against Δx , made non-dimensional by the correlation distance d_x	71
A.1	Example of function $Z(U_1, U_2)$	76
A.2	Linearisation of the Z -function (dark red plane) of the Z -function of Figure A.1 in a selected location (white dot).	76
A.3	Schematic view of the method of numerical integration for an example of two random variables. A 2-dimensional grid is defined and the Z -function is evaluated at the centre of the grid cells. Red grid points indicate failure ($Z < 0$), green grid point indicate no failure ($Z \geq 0$).	80
A.4	Schematic view of Monte Carlo sampling for an example with two random variables. Each dot represents a sampled pair (x_1, x_2) . Red dots indicate failure ($Z < 0$), green dots indicate no failure ($Z \geq 0$).	81

A.5	Relation between k and P_k according to equation (A.18). Function ϕ is the standard normal density function	82
A.6	Concept of importance sampling; left shows the concept of density shifting, right shows the concept of increased variance. The limit state function (LSF) is illustrated. Note that the LSF's are linear in this figure, but this is by no means a requirement for the applicability of importance sampling.	84
A.7	Schematic view of the directional sampling method in the standard normal space. The distances λ_i to the limit state ($Z = 0$) are shown for unit-length directional vectors θ_i for three directional samples ($i = 1, 2, 3$).	86
A.8	Method 1 to determine λ_i , the distance to the limit state $Z = 0$ for the i^{th} sampled direction.	87
A.9	Method 2 (gradient-based) to determine λ_i , the distance to the limit state $Z = 0$ for the i^{th} sampled direction.	88
A.10	Illustration of the design point in the U -space	89
A.11	Illustration of the FORM approximation	90
A.12	Contour lines of the example Z -function and the starting location (red dot) of the FORM procedure	94
A.13	Sampling the Z -function in all directions to estimate the derivative of Z to all u -variables.	95
A.14	Contour lines of the linearised Z -function	95
A.15	Contour lines of the normalised linearised Z -function	96
A.16	Estimated location of the design point based on the normalised linearised Z -function	96
A.17	Proces of relaxation. The red dot is the location of the current iteration step, the yellow dot is the estimated location of the design point based on the normalised linearised Z -function. The green dot shows the selected location of the next iteration, which is somewhere on the line between the red dot and the yellow dot.	97
A.18	Resulting steps in the iteration procedure.	97
A.19	Schematic view of the method centre of gravity.	99
A.20	Failure domain for a parallel system of two components – the shaded area indicates the combinations of Z -values that will lead to failure of the system.	102
A.21	The $Z_1 = 0$ and $Z_2 = 0$ contours in the U -space; hatched area indicates the area that contributes to the failure probability.	105
A.22	Replacement of " $Z_1 < 0 \cup Z_2 < 0$ " with " $Z^e < 0$ ". Failure domains of the three Z -functions are on the upper right side of the respective lines.	109
A.23	Functions Z_1 , Z_2 and Z_3 of the current example	114
A.24	Function Z_2 and the equivalent Z -function of $Z_1 \cup Z_3$	115
A.25	Functions $Z_1=0$, $Z_2 = 0$ and the equivalent Z -function of $Z_1 < 0 \cup Z_2 < 0$	116
A.26	Example of combining failure probabilities over four components	116
B.1	Procedure for determining a load variable associated with randomly selected standard normally distributed variables for the case of correlated variables.	117
B.2	Procedure for samples $u_{1,cor}$ and $u_{2,cor}$ of correlated standard uniform random variables $U_{1,cor}$ and $U_{2,cor}$	118
B.3	Uncorrelated (left) and correlated (right) samples according to the Gaussian correlation model assuming $\rho = 0.5$	120

1 Introduction

The Probabilistic Library provides functions for performing reliability and uncertainty analyses on any model, ranging from Python scripts to dedicated applications in the geotechnical or hydrodynamical fields, as well as combinations of models from various disciplines.

Reliability analysis determines the reliability – or probability of failure – of a physical structure. This analysis provides insight into the likelihood of an unwanted event occurring within a given period. For example, the probability of dike failure can be calculated, which is essential for dike assessments. Based on this probability, decisions can be made regarding whether reinforcement is necessary. Accurate failure probability calculations are crucial, as they help avoid unnecessary strengthening, thereby saving costs.

To improve the accuracy of failure probability calculations, survived events can be incorporated. Even with extensive efforts to measure input values, it is impossible to fully characterize the subsoil along long dike stretches. As a result, a model may predict dike failure, while real-world observations show that the dike has survived high water events. The Probabilistic Library integrates these observed events to refine and update failure probability estimates.

Another application of the Probabilistic Library is risk-based asset management. In this approach, a maintenance schedule is developed to minimize total costs over an extended period of time. These costs include both strengthening actions and risk-related expenses, where risk is defined as the probability of failure multiplied by the associated damage.

This document describes the scientific background of the Probabilistic Library **25.3.1**.

1.1 Overview of this document

[Chapter 2](#) covers distribution functions and correlation models available in the Probabilistic Library. [Chapter 3](#) and [chapter 4](#) present techniques for sensitivity analysis and uncertainty analysis, respectively. [Chapter 5](#) describes reliability analysis for both individual components and systems of components, along with the available reliability methods.

2 Statistics

Understanding probability distributions and their relationships is essential for statistical modeling and uncertainty analysis. This chapter covers distribution functions, which describe how values of a random variable are distributed, and correlations, which measure dependencies between variables.

To ensure that chosen distributions accurately represent data, we introduce a goodness-of-fit test, which assesses how well a theoretical distribution aligns with observed data. Additionally, we explore prior distributions, which play a crucial role in Bayesian analysis by incorporating prior knowledge into probabilistic models.

Together, these concepts provide a foundation for reliable statistical inference and risk assessment.

2.1 Distributions

2.1.1 Generic description of the application of distribution functions

The following three types of functions can be used to describe the statistical properties of random variables:

- 1 Cumulative distribution function (CDF);
- 2 Inverse cumulative distribution function (inverse CDF);
- 3 Probability density function (PDF).

The CDF, $F(x)$, provides the probability of non-exceedance, p , of each potential realisation, x , of random variable X . The inverse CDF, $F^{-1}(p)$, provides the realisation x that has a probability of non-exceedance p . The relation between the CDF and the inverse CDF is thus as follows:

$$F(x) = p \Leftrightarrow x = F^{-1}(p) \quad (2.1)$$

The PDF, $f(x)$, is the derivative of the CDF:

$$F(x) = \int_{-\infty}^x f(\tau) d\tau \Leftrightarrow f(x) = \frac{dF}{dx}(x) \quad (2.2)$$

The PDF provides the probability density for any given value of x . The probability density is the probability per unit value. For the normal distribution function, the pdf is the "famous" bell-shaped curve. As an example, Figure 2.1 shows the CDF, inverse CDF and PDF of the standard normal distribution function.

In discrete distributions the probability is only defined at a limited, countable number of x -values. In this case a PDF is not possible, because no derivative is available. Instead the PMF, the probability mass function is used. All PMF values sum up to 1.

The PMF, $f(x)$, relates as follows to the CDF:

$$F(x) = \sum_{-\infty}^x f(\tau) \quad (2.3)$$

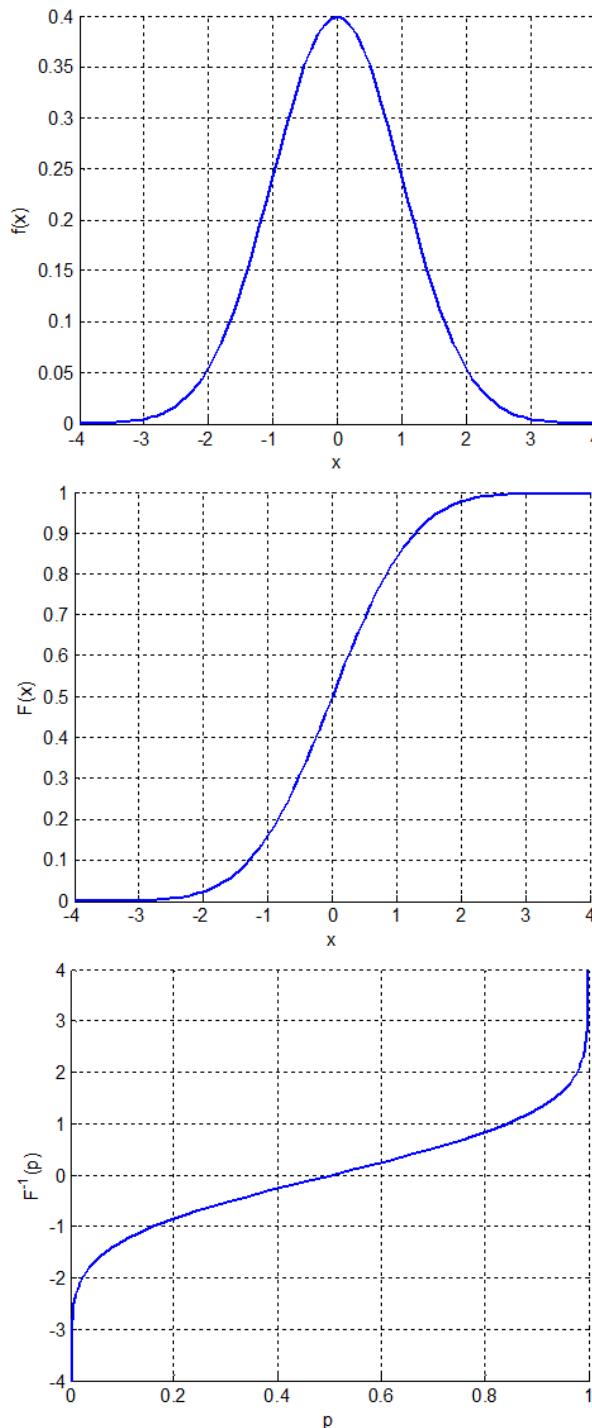


Figure 2.1: PDF (top), CDF (middle) and inverse CDF (bottom) of the standard normal distribution function.

The CDF, $F(x)$, has the following properties:

- 1 $F(x)$ is non-decreasing;
- 2 $\lim_{x \rightarrow -\infty} F(x) = 0$;
- 3 $\lim_{x \rightarrow \infty} F(x) = 1$.

Property 1 can be easily proven: if $x_1 < x_2$, then $P[X \leq x_1] \leq P[X \leq x_2]$ and thus

$F(x_1) \leq F(x_2)$. For a formal proof of properties 2 and 3, the reader is referred to [Grimmett and Stirzaker \(1983\)](#). But even without a proof it is intuitively clear that a realization from a probability function will be lower than ∞ and higher than $-\infty$.

Since $F(x)$ is non-decreasing, the inverse CDF, $F^{-1}(p)$, is also a non-decreasing function. In probabilistic computations, mainly the inverse CDF, $F^{-1}(x)$, of a variable X is applied, as schematically depicted in [Figure 2.2](#). The library of distribution functions therefore mostly consists of inverse CDF's. The procedure of [Figure 2.2](#) is explained below.

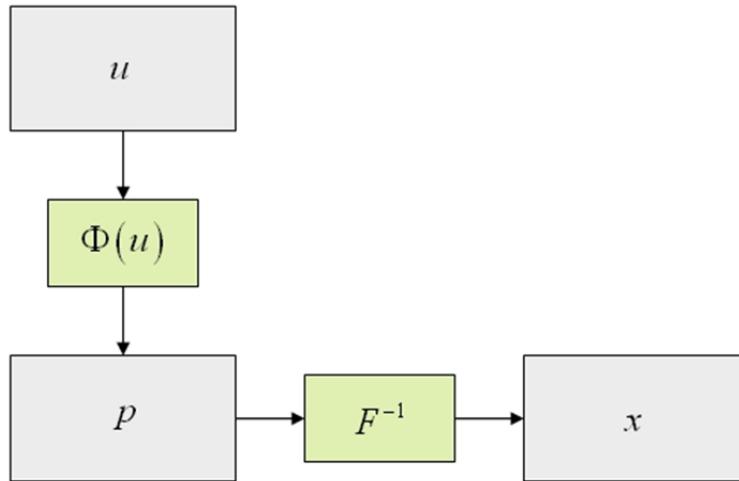


Figure 2.2: Procedure for determining a load variable associated with a randomly selected standard normally distributed variable (u -value), for the case of uncorrelated variables.

As described in section [section A.2](#), random variables are represented by standardised U -variables in the probabilistic computations in the Probabilistic Library, and the function $Z(U)$ is explored to derive an estimate of the failure probability. In order to evaluate function $Z(U)$, the realisations of the U -variables are first translated to the corresponding realisations of the X -variables and subsequently the Z -value is determined. Assume for the sake of simplicity that the X -variables are mutually independent (correlations will be dealt with in [section 2.2](#)). As explained in [section 2.3](#), the transformation from a realization, u , of variable U , to realization x , of variable X , is done in such a way that the (non-)exceedance probabilities of u and x are equal. This transformation, as depicted in [Figure 2.2](#), can be formulated as follows:

$$\Phi(u) = F(x) \quad \Rightarrow \quad x = F^{-1}(\Phi(u)) \quad (2.4)$$

where:

- Φ is the standard normal distribution function
- F is the CDF of variable X
- F^{-1} is the inverse CDF of X
- x is a realisation of X
- u is a realisation of U

This procedure automatically guarantees that variable x is a realization from distribution function $F(x)$ and therefore correctly represents the statistical properties of variable X . This is demonstrated below.

First it needs to be shown that the value $p = \Phi(u)$ is a realization from a standard uniform distribution function. The standard uniform distribution function is the CDF in which each value in the range $[0, 1]$ has equal probability density. The CDF of this function is as follows (see also [Figure 2.3](#)):

$$F(x) = \begin{cases} 0 & ; x \leq 0 \\ x & ; 0 < x < 1 \\ 1 & ; x \geq 1 \end{cases} \quad (2.5)$$

Consider a realization, u^* , of the standard normal distribution function with a probability of non-exceedance equal to $p^* = \Phi(u^*)$. By definition this means that the probability that a random sample u from the standard normal distribution function does not exceed u^* is equal to p^* . In formula:

$$P[u \leq u^*] = p^* \quad (2.6)$$

Since Φ is a CDF, it is a non-decreasing function. Therefore it follows from equation (2.6) that:

$$P[\Phi(u) \leq \Phi(u^*)] = p^* \quad (2.7)$$

And since by definition $p^* = \Phi(u^*)$, this simplifies to:

$$P[\Phi(u) \leq p^*] = p^* \quad (2.8)$$

So the probability that $\Phi(u)$ does not exceed a given value p^* ($0 \leq p^* \leq 1$) is equal to p^* . This shows that $\Phi(u)$ is a realization from a standard uniform distribution function, as described by equation (2.5) and depicted in [Figure 2.3](#).

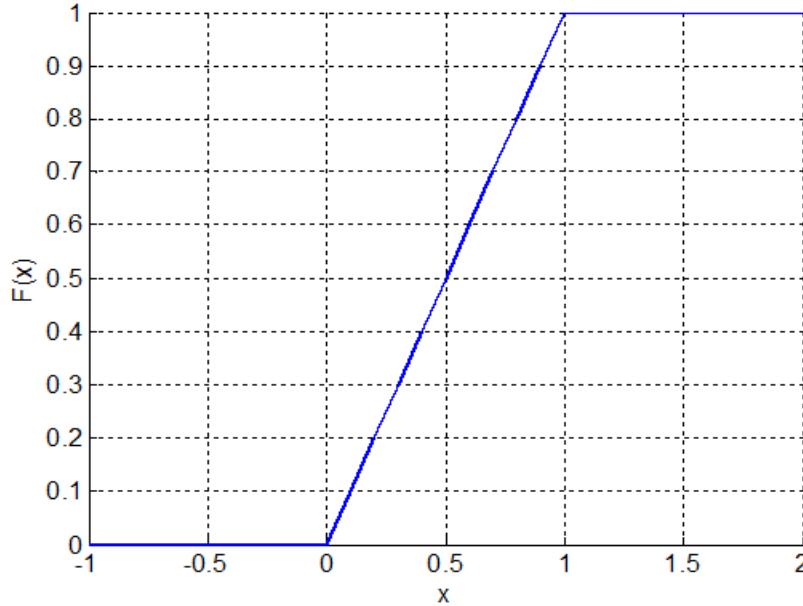


Figure 2.3: Standard uniform distribution function (CDF).

It has been demonstrated that the value p in [Figure 2.2](#) is a realization from the standard uniform distribution function. The next step is to show that the value $x = F^{-1}(p)$ in [Figure 2.2](#)

is a realization from distribution function $F(x)$. For this purpose, consider a value x^* with probability of non-exceedance p^* . This means: $F(x^*) = p^*$ and $x^* = F^{-1}(p^*)$. The value p in Figure 2.2 is taken from a standard uniform distribution function (as proven above), which means:

$$P[p \leq p^*] = p^* \quad (2.9)$$

Since F is an inverse CDF, it is a non-decreasing function and therefore it follows from equation (2.9) that:

$$P[F^{-1}(p) \leq F^{-1}(p^*)] = p^* \quad (2.10)$$

By definition $x^* = F^{-1}(p^*)$ and $x = F^{-1}(p)$, which means equation (2.10) simplifies to:

$$P[x \leq x^*] = p^* \quad (2.11)$$

Since $p^* = F(x)$, this means:

$$P[x \leq x^*] = F(x^*) \quad (2.12)$$

This shows that value x in Figure 2.2 is a realization from distribution function $F(x)$.



Note: the Probabilistic Library works with standard normalized U -variables (see section 2.3). The limit state function $Z(U)$ is explored to derive an estimate of the failure probability. In order to evaluate function $Z(U)$, the realisations of the U -variables are first translated to the corresponding realisations of the X -variables to be able to determine the Z -value. In this transformation, the inverse CDF of variable X is applied, to provide variable X with the correct statistical properties.

2.1.2 Distribution properties

In theory, endless distribution types exist. Each distribution reflects actual values. From these values, x_i , the following properties can be derived:

- ◊ Mean or expectation value μ : The long run average of randomly chosen values, calculated as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.13)$$

- ◊ Mode: The value with the highest probability density;
- ◊ Median: The value for which half of the randomly chosen values is less than this value and half of the randomly chosen values is more than this value;
- ◊ Standard deviation σ : A measure how much randomly chosen values differ from the mean, calculated as follows:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.14)$$

- ◊ Variation coefficient V : Relative deviation with respect to the mean:

$$V = \frac{\sigma}{\mu} \quad (2.15)$$

- ◊ Quantile value Q : Generalization of the median. From all randomly chosen values, a user provided fraction is less than this value and the remainder is more than this value.

2.1.3 Distribution types

This paragraph lists all distribution types supported by Probabilistic Library. The distribution functions often require the following parameters as inputs (depending on the distribution type):

- ◊ Location m : An indication where the distribution is located. For some distributions the mean μ is equal to the location m .
- ◊ Scale s : An indication how much randomly chosen values differ from the location. For some distributions the scale s is equal to the deviation σ .
- ◊ Shape k : Describes the shape of the distribution.
- ◊ Shift c : The distribution is shifted a certain amount. Not present in all distributions.
- ◊ Minimum and Maximum a and b : Minimum and maximum possible values of randomly chosen values. Not present in all distributions.
- ◊ Number of observations N : Indicates the number of values when fitted from a collection with values x_i .
- ◊ Prior distribution prior: Used when fitted from data and a prior distribution. The prior distribution often refers to an overall or long term known distribution. It is used to derive a distribution for a specific occasion, where limited data is available. Only applicable for a few distribution types.

2.1.3.1 Deterministic distribution

The deterministic distribution function is defined as follows:

$$\text{PMF} \quad f(x) = \begin{cases} x \neq m & 0 \\ x = m & 1 \end{cases}$$

$$\text{CDF} \quad F(x) = \begin{cases} x < m & 0 \\ x \geq m & 1 \end{cases}$$

$$\text{Mean} \quad \mu = m$$

$$\text{Deviation} \quad \sigma = 0$$

$$\text{Fit} \quad m = \frac{1}{N} \sum_{i=1}^N x_i$$

2.1.3.2 Normal distribution

The normal distribution function is defined as follows:

PDF	$f(x) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2s^2}\right)$
CDF	$F(x) = \Phi\left(\frac{x-m}{s}\right)$
Mean	$\mu = m$
Deviation	$\sigma = s$
Fit	$m = \frac{1}{N} \sum_{i=1}^N x_i$
	$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - m)^2$
Fit with prior	$\frac{1}{s^2} = \frac{1}{s_{\text{prior}}^2} + \frac{N}{s_{\text{fit}}^2}$ $m = s \cdot \left(\frac{m_{\text{prior}}}{s_{\text{prior}}^2} + N \cdot \frac{m_{\text{fit}}}{s_{\text{fit}}^2} \right)$

Figure 2.4 shows the probability density of the normal distribution, with the parameters μ and σ indicated. Figure 2.5 shows the variation in the density function for different choices of μ and σ .

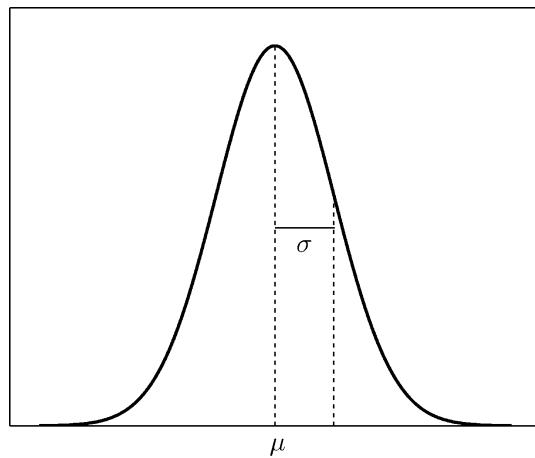


Figure 2.4: Normal probability density function, with parameters μ and σ indicated

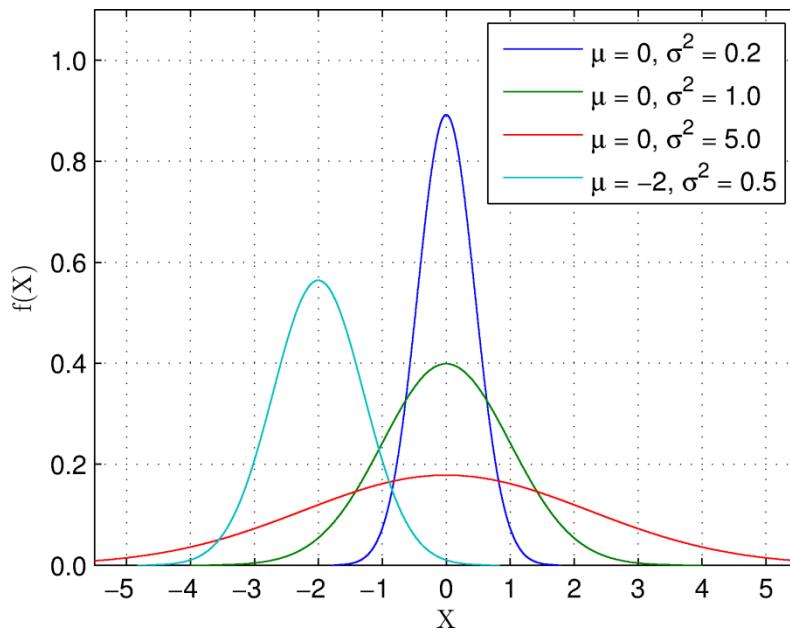


Figure 2.5: Illustration of the effect of parameters μ and σ

2.1.3.3 Standard normal distribution

The standard normal distribution is a normal distribution with mean $\mu = 0$ and deviation $\sigma = 1$. The standard normal distribution is unlikely to occur in the real world, but is used internally in the Probabilistic Library.

The corresponding probability density function is:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad (2.16)$$

and the cumulative distribution function, or in other words the non exceeding probability, is:

$$\Phi(u) = \int_{-\infty}^u \varphi(v) dv \quad (2.17)$$

Since there is no closed form to express Φ , it is approximated by approximation formula 26.2.17 for Normal Probability Function, Handbook of Mathematical Functions, Abramowitz & Stegun. In Probabilistic Library this is done using the methods of [Wichura \(1988\)](#) and [Hart et al. \(1968\)](#). See [Appendix C](#) for more details.

Other distribution types are converted to the standard normal distribution. The physical value in another distribution type, called x , is converted to a value u in the standard normal distribution, in such a way that the non exceeding probability of x is equal to the exceeding probability of u . With Φ the cumulative density function of the standard normal distribution, the converted value u is:

$$u(x) = \Phi^{-1}(F(x)) \quad (2.18)$$

The probabilities of the standard normal distribution are displayed in the following figure.

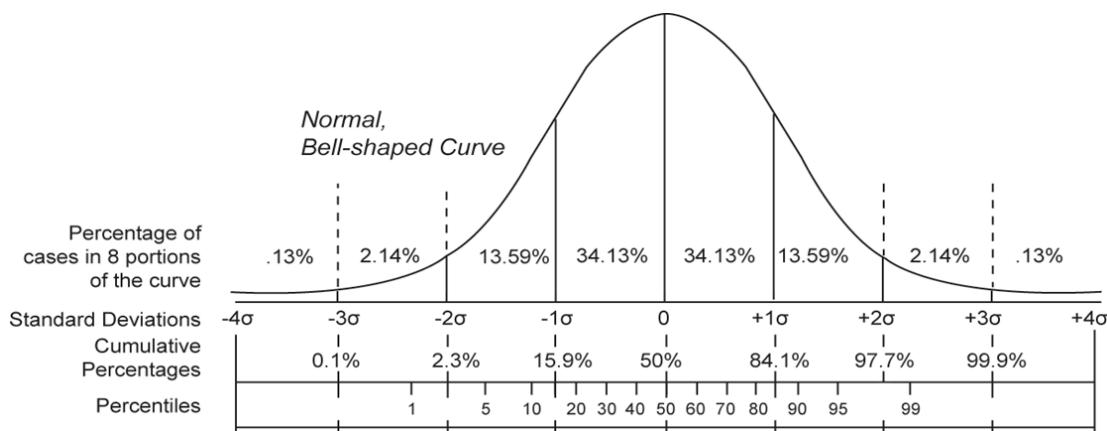


Figure 2.6: Standard normal distribution

2.1.3.4 Log normal distribution

If parameter $y = \ln(x)$ has a normal distribution, then parameter x has a log-normal distribution. A log-normal distribution always yields values higher than a given shift (usually 0). The normal and log-normal distributions are similar for small ratios between the standard deviation and the mean.

The log normal distribution function is defined as follows:

$$\begin{aligned}
 \text{PDF} \quad f(x) &= \frac{1}{(x-c)s\sqrt{2\pi}} \exp\left(-\frac{(\ln(x-c)-m)^2}{2s^2}\right) \\
 \text{CDF} \quad F(x) &= \Phi\left(\frac{\ln(x-c)-m}{s}\right) \\
 \text{Mean} \quad \mu &= \exp(m + \frac{1}{2}s^2) \iff m = \ln(\mu) - \frac{1}{2}s^2 \\
 \text{Deviation} \quad \sigma^2 &= (\mu - c)^2 \cdot (\exp(s^2) - 1) \iff s^2 = \ln\left(1 + \left(\frac{\sigma}{\mu}\right)^2\right) \\
 \text{Fit} \quad m &= \frac{1}{N} \sum_{i=1}^N \ln(x_i - c) \\
 &\quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (\ln(x_i - c) - m)^2 \\
 \text{Fit with prior} \quad \frac{1}{s^2} &= \frac{1}{s_{\text{prior}}^2} + \frac{N}{s_{\text{fit}}^2} \\
 &\quad m = s \cdot \left(\frac{m_{\text{prior}}}{s_{\text{prior}}^2} + N \cdot \frac{m_{\text{fit}}}{s_{\text{fit}}^2} \right)
 \end{aligned}$$

2.1.3.5 Uniform distribution

The uniform distribution function is defined as follows:

$$\begin{aligned} \text{PDF} \quad f(x) &= \begin{cases} 0 & x < a \vee x > b \\ \frac{1}{b-a} & x \geq a \wedge x \leq b \end{cases} \\ \text{CDF} \quad F(x) &= \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \geq a \wedge x \leq b \\ 1 & x > b \end{cases} \\ \text{Mean} \quad \mu &= \frac{1}{2}(a+b) \\ \text{Deviation} \quad \sigma^2 &= \frac{1}{12}(b-a)^2 \\ \text{Fit} \quad a &= x_{\min} - \delta \\ &b = x_{\max} + \delta \\ \text{with} \quad \delta &= \frac{x_{\max} - x_{\min}}{N} \end{aligned}$$

The distribution parameters a and b indicate the range over which the probability density function is non-zero, with a indicating the starting point and b indicating the ending point. [Figure 2.7](#) and [Figure 2.8](#) show the uniform probability density and cumulative probability distribution, respectively, as a function of a and b .

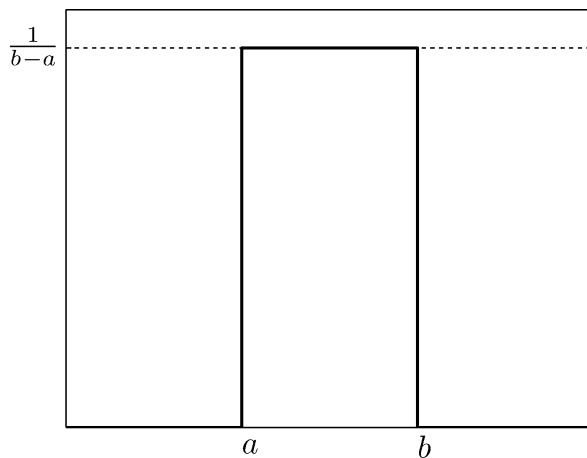


Figure 2.7: Uniform probability density function, with parameters a and b indicated

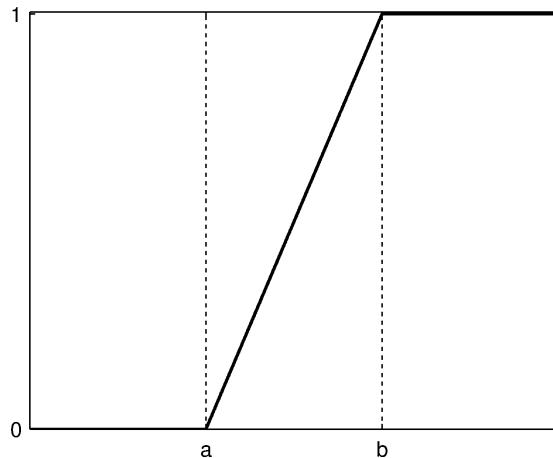


Figure 2.8: Uniform cumulative distribution function, with parameters a and b indicated

2.1.3.6 Triangular distribution

The properties of the triangular distribution are defined by a minimum a , a maximum b and a shift c . The mode of the distribution is equal to c . It is required that $a < c < b$.

$$\text{PDF} \quad f(x) = \begin{cases} 0 & x < a \vee x > b \\ \frac{2(x-a)}{(b-a)(c-a)} & x \geq a \wedge x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & x \geq c \wedge x \leq b \\ 0 & x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & x \geq a \wedge x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & x \geq c \wedge x \leq b \\ 1 & x > b \end{cases}$$

$$\text{Mean} \quad \mu = \frac{1}{3}(a + b + c)$$

$$\text{Deviation} \quad \sigma^2 = \frac{1}{18}(a^2 + b^2 + c^2 - ab - ac - bc)$$

$$\text{Fit} \quad a = x_{\min} - \delta$$

$$b = x_{\max} + \delta$$

$$c = 3\mu_x - (a + b)$$

with

$$\delta = \frac{x_{\max} - x_{\min}}{N}$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$$

2.1.3.7 Trapezoidal distribution

The properties of the trapezoidal distribution are defined by a minimum a , a maximum b , lower shift c and upper shift d . The mode of the distribution is between c and d . It is required that $a < c < d < b$. The effective length L is the mean of the difference between the maximum and minimum and the difference between the upper shift and lower shift.

$$\begin{aligned} \text{Length} \quad L &= \frac{b - a + d - c}{2} \\ \text{PDF} \quad f(x) &= \begin{cases} 0 & x < a \vee x > b \\ \frac{x-a}{L(c-a)} & x \geq a \wedge x \leq c \\ \frac{1}{L} & x \geq c \wedge x \leq d \\ \frac{b-x}{L(b-d)} & x \geq d \wedge x \leq b \end{cases} \\ \text{CDF} \quad F(x) &= \begin{cases} 0 & x < a \\ \frac{(x-a)^2}{2L(b-a)} & x \geq a \wedge x \leq c \\ \frac{2x-a-c}{2L} & x \geq c \wedge x \leq d \\ 1 - \frac{(b-x)^2}{2L(b-d)} & x \geq d \wedge x \leq b \\ 1 & x > b \end{cases} \\ \text{Mean} \quad \mu &= \frac{1}{6L} \left(\frac{b^3 - d^3}{b - d} - \frac{c^3 - a^3}{c - a} \right) \\ \text{Deviation} \quad \sigma^2 &= \frac{1}{12L} \left(\frac{b^4 - d^4}{b - d} - \frac{c^4 - a^4}{c - a} \right) - \mu^2 \end{aligned}$$

2.1.3.8 Exponential distribution

The exponential distribution is defined as follows:

$$\begin{aligned} \text{PDF} \quad f(x) &= \frac{1}{s} \exp\left(-\frac{x-c}{s}\right) \\ \text{CDF} \quad F(x) &= 1 - \exp\left(-\frac{x-c}{s}\right) \\ \text{Mean} \quad \mu &= s + c \\ \text{Deviation} \quad \sigma &= s \\ \text{Rate} \quad \lambda &= \frac{1}{s} \\ \text{Fit} \quad s &= \frac{1}{N} \sum_{i=1}^N (x_i - c) \end{aligned}$$

The parameter of the exponential distribution, λ , is referred to as the rate parameter, and determines how quickly the density function goes to zero. The parameter c is the shift parameter and serves to horizontally shift the density.

2.1.3.9 Gumbel distribution

The Gumbel distribution function is defined as follows, where s is the scale parameter and c is the shift parameter. The shift parameter is also called the location parameter.

$$\begin{aligned}
 \text{PDF} \quad f(x) &= \frac{1}{s} \exp \left(- \left(x + \exp \left(-\frac{x-c}{s} \right) \right) \right) \\
 \text{CDF} \quad F(x) &= \exp \left(\exp \left(-\frac{x-c}{s} \right) \right) \\
 \text{Mean} \quad \mu &= c + s \cdot \gamma \\
 \text{Deviation} \quad \sigma^2 &= \frac{\pi^2}{6} s^2 \\
 \text{Fit} \quad s &= \frac{1}{N} \sum_{i=1}^N x_i + \frac{\sum_{i=1}^N x_i \exp \left(-\frac{x_i}{s} \right)}{\sum_{i=1}^N \exp \left(-\frac{x_i}{s} \right)} \\
 c &= -s \cdot \ln \left(\frac{1}{N} \sum_{i=1}^N \exp \left(-\frac{x_i}{s} \right) \right)
 \end{aligned}$$

where:

γ is the Euler-Mascheroni constant (0.5772156649...)

The expression for the fit originates from ([Forbes, 2010](#)).

Decimation value

Alternatively, the location and scale can be derived from a decimation value D and a known exceeding probability for a certain x_{exc} . The decimation value D is the difference in the x -value, which reduces the exceeding probability F_{exc} with a factor 10. This leads to:

$$F_{\text{exc}}(x_{\text{exc}} + D) = \frac{1}{10} F_{\text{exc}}(x_{\text{exc}}) \quad (2.19)$$

with:

$$F_{\text{exc}}(x) = 1 - F(x) \quad (2.20)$$

The distribution properties can be derived as follows from the decimation height:

$$\begin{aligned}
 \text{Fit from } D \quad s &= \frac{D}{z(F_{\text{exc}}) - z(\frac{1}{10} F_{\text{exc}})} \approx \frac{D}{\ln(10)} \\
 m &= x_{\text{exc}} + s \cdot z(F_{\text{exc}}(x_{\text{exc}})) \\
 \text{with} \\
 z(F) &= \ln(-\ln(1 - F))
 \end{aligned}$$

2.1.3.10 Weibull distribution

The Weibull distribution is defined by the scale parameter s , the shape parameter k and the shift parameter c . The shift parameter is also called the location parameter. The Weibull distribution is described as follows:

$$\begin{aligned} \text{PDF} \quad f(x) &= \frac{k}{s} \left(\frac{x-c}{s} \right)^{k-1} \exp \left(-\left(\frac{x-c}{s} \right)^k \right) \\ \text{CDF} \quad F(x) &= 1 - \exp \left(-\left(\frac{x-c}{s} \right)^k \right) \\ \text{Mean} \quad \mu &= c + s \cdot \Gamma \left(1 + \frac{1}{k} \right) \\ \text{Deviation} \quad \sigma^2 &= s^2 \cdot \Gamma \left(1 + \frac{2}{k} \right) - \mu^2 \\ \text{Fit} \quad k &= \frac{1}{k} + \frac{\sum_{i=1}^N \ln(x_i)}{N} - \frac{\sum_{i=1}^N x_i^k \cdot \ln(x_i)}{\sum_{i=1}^N x_i^k} \\ s^k &= \frac{\sum_{i=1}^N x_i^k}{N} \end{aligned}$$

where:

Γ is the gamma function;

The expression for the fit originates from (García, 1981).

2.1.3.11 Conditional Weibull distribution

The conditional Weibull distribution gives the probability that $X = x$ conditionally on the event that $X > \omega$, where ω represents a threshold. It is used in when considering a peaks-over-threshold (POT) method. The distribution function of the conditional Weibull is given as follows:

$$P(X \leq x | X > \omega) = 1 - \exp \left[(\omega/\sigma)^\xi - (x/\sigma)^\xi \right] \quad (2.21)$$

where the parameters ω , σ , and ξ refer to the threshold, scale, and shape parameter, respectively. The conditional Weibull distribution is often described in terms of exceedance frequencies rather than probabilities. The exceedance frequency of x can be described as follows:

$$Fr(X > x) = \lambda \cdot P(X > x | X > \omega) \quad (2.22)$$

where Fr refers to ‘frequency’, and λ is the frequency with which the selected threshold ω is exceeded:

$$\lambda = Fr(X > \omega) \quad (2.23)$$

In practice, λ is determined by counting the number of independent peaks above the threshold and dividing by the number of years of record.

Expanding equation (2.22) so that the probability is full written out gives the following form of the exceedance frequency distribution for the condition Weibull:

$$Fr(X > x) = \lambda \exp \left[(\omega/\sigma)^\xi - (x/\sigma)^\xi \right] \quad (2.24)$$

2.1.3.12 Frechet distribution

The Frechet distribution is defined by the scale parameter s , the shape parameter k and the shift parameter c . The shift parameter is also called the location parameter. The Frechet properties are:

PDF	$f(x) = \frac{k}{s} \left(\frac{x-c}{s} \right)^{-1-k} \exp \left(- \left(\frac{x-c}{s} \right)^{-k} \right)$
CDF	$F(x) = \exp \left(- \left(\frac{x-c}{s} \right)^{-k} \right)$
Mean	$\mu = s \cdot \Gamma \left(1 - \frac{1}{k} \right) + c$
Deviation	$\sigma^2 = s^2 \cdot \left(\Gamma \left(1 - \frac{2}{k} \right) - \Gamma^2 \left(1 - \frac{1}{k} \right) \right)$

where:

Γ is the gamma function

2.1.3.13 Generalized Extreme Value distribution

The Generalized Extreme Value distribution is a combination of the Gumbel, Frechet and inverted Weibull distribution. Depending on the sign of the shape parameter k , one of these distribution is used.

The Generalized Extreme Value distribution D_{GEV} is defined as follows:

$$D_{GEV}(s, k, c) = \begin{cases} k = 0 & D_{Gumbel}(s, c) \\ k > 0 & D_{Frechet}\left(\frac{s}{k}, \frac{1}{k}, c - \frac{s}{k}\right) \\ k < 0 & D_{Weibull, \text{inverted}}\left(-\frac{s}{k}, -\frac{1}{k}, c - \frac{s}{k}\right) \end{cases} \quad (2.25)$$

2.1.3.14 Rayleigh distribution

The Rayleigh distribution is defined as follows:

$$\begin{aligned}
 \text{PDF} \quad f(x) &= \frac{x - c}{s^2} \exp\left(-\frac{(x - c)^2}{2s^2}\right) \\
 \text{CDF} \quad F(x) &= 1 - \exp\left(-\frac{(x - c)^2}{2s^2}\right) \\
 \text{Mean} \quad \mu &= \sqrt{\frac{\pi}{2}} \cdot s + c \\
 \text{Deviation} \quad \sigma^2 &= \frac{4 - \pi}{2} \cdot s^2 \\
 \text{Fit} \quad c &= x_{\min} - \delta \\
 &\quad s^2 = \frac{1}{2N} \sum_{i=1}^N (x_i - c)^2 \\
 &\quad \text{with} \\
 &\quad \delta = \frac{x_{\max} - x_{\min}}{N}
 \end{aligned}$$

2.1.3.15 Rayleigh- N distribution

The Rayleigh- N distribution is a special distribution implemented in Probabilistic Library. This distribution is based on the Rayleigh distribution (see section 2.1.3.14) and it is defined as follows:

$$F(x; \sigma, N) = \left(1 - e^{-x^2/(2\sigma^2)}\right)^N \text{ for } x \geq 0 \quad (2.26)$$

where $\sigma > 0$ is the scale parameter and $N > 0$.

The Rayleigh- N distribution is in fact the Rayleigh distribution taken to the power N . When $N = 1$, the Rayleigh- N distribution is exactly equal to the Rayleigh distribution. The corresponding probability density function is:

$$f(x; \sigma, N) = \frac{Nx}{\sigma^2} \left(1 - e^{-x^2/(2\sigma^2)}\right)^{N-1} e^{-x^2/(2\sigma^2)} \text{ for } x \geq 0 \quad (2.27)$$

2.1.3.16 Pareto distribution

The Pareto distribution is defined as follows:

$$\begin{aligned}
 \text{PDF} \quad f(x) &= k \cdot \frac{s^k}{x^{k+1}} \\
 \text{CDF} \quad F(x) &= 1 - \left(\frac{s}{x}\right)^k \\
 \text{Mean} \quad \mu &= \frac{k \cdot s}{k - 1} \\
 \text{Deviation} \quad \sigma^2 &= \frac{k \cdot s^2}{(k - 1)^2 \cdot (k - 2)} \\
 \text{Fit} \quad s &= x_{\min} \\
 &\quad \frac{1}{k} = \frac{1}{N} \sum_{i=1}^N (\ln(x_i) - \ln(s))
 \end{aligned}$$

2.1.3.17 Generalized Pareto distribution

The Generalized Pareto distribution is defined as follows:

$$\begin{aligned} \text{PDF} \quad f(x) &= \frac{1}{2} \cdot \left(1 + k \left(\frac{x-m}{s}\right)\right)^{-\left(\frac{1}{k}+1\right)} \\ \text{CDF} \quad F(x) &= 1 - \left(1 + \left(\frac{x-m}{s}\right)\right)^{-\frac{1}{k}} \\ \text{Mean} \quad \mu &= m + \frac{s}{1-k} \\ \text{Deviation} \quad \sigma^2 &= \frac{s^2}{(1-k)^2 \cdot (1-2k)} \end{aligned}$$

2.1.3.18 Student's T distribution

The student's T distribution is defined as follows. It is based on the degrees of freedom ν , which is equal to $N - 1$.

$$\begin{aligned} \text{PDF} \quad f(x) &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\left(\frac{x-m}{s}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\ \text{CDF} \quad F(x) &= 1 - \frac{1}{2} I_{t(x)}\left(\frac{\nu}{2}, \frac{1}{2}\right) \\ &\text{with} \\ t(x) &= \frac{\nu}{\left(\frac{x-m}{s}\right)^2 + \nu} \\ \text{Mean} \quad \mu &= m \\ \text{Deviation} \quad \sigma^2 &= \frac{\nu}{\nu-2} s^2 \\ \text{Fit} \quad m &= \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} \\ s^2 &= \frac{\sum_{i=1}^N w_i (x_i - m)^2}{\nu + 1} \\ &\text{with} \\ w_i &= \frac{(\nu + 1) s^2}{\nu s^2 + (x_i - m)^2} \end{aligned}$$

The expression for the fit originates from
[\(Probabilityislogic , https://stats.stackexchange.com/users/2392/probabilityislogic\).](https://stats.stackexchange.com/users/2392/probabilityislogic)

2.1.3.19 Gamma distribution

The Gamma distribution function is described as follows:

$$\text{PDF} \quad f(x) = \frac{1}{\Gamma(k)} s^k x^{k-1} \exp\left(-\frac{x}{s}\right)$$

$$\text{CDF} \quad F(x) = 1 - \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{s}\right)$$

$$\text{Mean} \quad \mu = k \cdot s$$

$$\text{Deviation} \quad \sigma^2 = k \cdot s^2$$

$$\text{Fit} \quad k = \frac{3 - z + \sqrt{(3 - z)^2 + 24z}}{12z}$$

$$s = \frac{1}{kN} \sum_{i=1}^N x_i$$

with

$$z = \frac{1}{N} \ln \left(\sum_{i=1}^N x_i \right) - \frac{1}{N} \sum_{i=1}^N \ln(x_i)$$

where:

$\Gamma(\alpha)$ is the gamma function

$\gamma(\alpha, \beta)$ is the lower incomplete gamma function

2.1.3.20 Beta distribution

The Beta distribution is defined with two shape parameters, k_1 and k_2 , as follows:

$$\text{PDF} \quad f(x) = \frac{x^{k_1-1} (1-x)^{k_2-1}}{B(k_1, k_2)}$$

$$\text{CDF} \quad F(x) = I_x(k_1, k_2)$$

$$\text{Mean} \quad \mu = \frac{k_1}{k_1 + k_2}$$

$$\text{Deviation} \quad \sigma^2 = \frac{k_1 k_2}{(k_1 + k_2)^2 (k_1 + k_2 + 1)}$$

$$\text{Fit} \quad k_1 = \left(\frac{1-m}{s^2} - \frac{1}{m} \right) m^2$$

$$k_2 = k_1 \left(\frac{1}{m} - 1 \right)$$

with

$$m = \frac{1}{N} \sum_{i=1}^N x_i$$

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - m)^2$$

where:

$B(\alpha, \beta)$ is the beta function;

$I_x(\alpha, \beta)$ is the regularized incomplete beta function;

2.1.3.21 Poisson distribution

The Poisson distribution is defined as follows (where $\lfloor x \rfloor$ is the floor function of x):

$$\text{PMF} \quad f(x) = \begin{cases} x \in \mathbb{N} & \exp(-m) \cdot \frac{m^x}{x!} \\ x \notin \mathbb{N} & 0 \end{cases}$$

$$\text{CDF} \quad F(x) = \exp(-m) \cdot \sum_{i=0}^{\lfloor x \rfloor} \frac{m^i}{i!}$$

$$\text{Mean} \quad \mu = m$$

$$\text{Deviation} \quad \sigma^2 = m$$

2.1.3.22 Discrete distribution

The properties of the discrete distribution are based on a set of values $\{x_1, x_2, \dots, x_N\}$ and their probabilities of occurrence w_i :

$$\text{PMF} \quad f(x) = \begin{cases} x = x_i & \frac{w_i}{\sum_{j=1}^N w_j} \\ x \neq x_i & 0 \end{cases}$$

$$\text{CDF} \quad F(x) = \sum_{x_i \leq x} f(x)$$

$$\text{Mean} \quad \mu = \frac{\sum_{i=1}^N w_i \cdot x_i}{\sum_{i=1}^N w_i}$$

$$\text{Deviation} \quad \sigma^2 = \frac{\sum_{i=1}^N w_i \cdot (x_i - \mu)^2}{\sum_{i=1}^N w_i}$$

2.1.3.23 Bernoulli distribution

The Bernoulli distribution is a special case of a discrete distribution. It consists of two discrete values, one at 0 with probability $1 - m$ and one at 1 with probability m .

$$\begin{aligned}
 \text{PMF} \quad f(x) &= \begin{cases} x = 0 & 1 - \mu \\ x = 1 & \mu \\ x \notin \{0, 1\} & 0 \end{cases} \\
 \text{CDF} \quad F(x) &= \begin{cases} x < 0 & 0 \\ 0 \leq x \leq 1 & 1 - \mu \\ x > 1 & 1 \end{cases} \\
 \text{Mean} \quad \mu &= m \\
 \text{Deviation} \quad \sigma^2 &= m \cdot (1 - m) \\
 \text{Fit} \quad m &= \frac{1}{N} \sum_{i=1}^N x_i \\
 \text{Fit with prior} \quad m &= \frac{N_{\text{prior}} \cdot m_{\text{prior}} + N_{\text{fit}} \cdot m_{\text{fit}}}{N_{\text{prior}} + N_{\text{fit}}}
 \end{aligned}$$

2.1.3.24 CDF curve distribution

The CDF curve distribution consists of a number of reliability indices for which the physical value is defined. The CDF curve is very much like a fragility curve, only the CDF curve requires that it is strictly monotone ascending or descending.

2.1.3.25 Histogram distribution

The histogram distribution consists of a number of bins. Within each bin, which is defined with a lower bound and an upper bound, each value is assumed to have the same probability of occurrence.

When fitted, the aim is to generate 100 non empty bins, all with the same width. The lowest bin has a lower boundary a and the highest bin has an upper boundary b .

If the lowest fitted value appears multiple times (unrounded), it is assumed that this value is a lower limit and no values lower than this value are possible. A bin with width zero and boundaries equal to the lowest value is added to the histogram. The same is applied for the highest model result.

$$\begin{aligned}
 \text{Fit} \quad a &= \begin{cases} N(x_{\min}) = 1 & x_{\min} - \delta \\ N(x_{\min}) > 1 & x_{\min} \end{cases} \\
 b &= \begin{cases} N(x_{\max}) = 1 & x_{\min} + \delta \\ N(x_{\max}) > 1 & x_{\max} \end{cases}
 \end{aligned}$$

where:

- $N(x_{\min})$ is the number of appearances of value x_{\min}
- $N(x_{\max})$ is the number of appearances of value x_{\max}
- $\delta = \frac{x_{\max} - x_{\min}}{N}$

2.1.3.26 Inverted distribution

The inverted distribution is a distribution on top of another base distribution. The inverted distribution is "mirrored" in the x-direction around the shift value c_{base} of the base distribution, or 0 if the base distribution does not have a shift.

The inverted distribution is defined as follows:

$$\begin{aligned}\text{PDF} \quad f(x) &= f_{\text{base}}(c_{\text{base}} - x) \\ \text{CDF} \quad F(x) &= 1 - F_{\text{base}}(c_{\text{base}} - x) \\ \text{Mean} \quad \mu &= 2c_{\text{base}} - \mu_{\text{base}} \\ \text{Deviation} \quad \sigma &= \sigma_{\text{base}}\end{aligned}$$

2.1.3.27 Truncated distribution

The inverted distribution is a distribution on top of another base distribution and is truncated below a minimum value a and maximum value b . The mean and standard deviation displayed are the properties of the base distribution, although not being the correct mean and standard deviation. The truncated distribution is

$$\begin{aligned}\text{PDF} \quad f(x) &= \begin{cases} x < a \vee x > b & 0 \\ x \geq a \wedge x \leq b & \frac{f_{\text{base}}(x)}{F_{\text{base}}(b) - F_{\text{base}}(a)} \end{cases} \\ \text{CDF} \quad F(x) &= \begin{cases} x < a & 0 \\ x \geq a \wedge x \leq b & \frac{F_{\text{base}}(x) - F_{\text{base}}(a)}{F_{\text{base}}(b) - F_{\text{base}}(a)} \\ x > b & 1 \end{cases} \\ \text{Mean} \quad \mu &= \mu_{\text{base}} \\ \text{Deviation} \quad \sigma &= \sigma_{\text{base}}\end{aligned}$$

2.1.4 Prior distribution

A prior distribution is used to update another distribution, resulting in a posterior distribution. The prior distribution should represent a general distribution of a parameter, such as a worldwide distribution or a long-term distribution. This distribution can be updated using local data or data that is only applicable to a specific situation. The resulting updated distribution is then applicable to that situation. The advantage of this approach is that a meaningful distribution can be derived even with a limited number of measurements.

Fitting from a posterior distribution is only possible for normal distributions. The updated posterior distribution is calculated as follows [Lynch (2007)]:

$$\mu_{\text{post}} = \frac{\sigma_{\text{data}}^2 \cdot \mu_{\text{prior}} + n \cdot \sigma_{\text{prior}}^2 \cdot \mu_{\text{data}}}{n \cdot \sigma_{\text{prior}}^2 + \sigma_{\text{data}}^2} \quad (2.28)$$

and

$$\sigma_{\text{post}}^2 = \frac{\sigma_{\text{data}}^2 \cdot \sigma_{\text{prior}}^2}{n \cdot \sigma_{\text{prior}}^2 + \sigma_{\text{data}}^2} \quad (2.29)$$

where:

- $\mu_{\text{post}}, \sigma_{\text{post}}$ are the mean and standard deviation of the updated (or posterior) distribution
- $\mu_{\text{prior}}, \sigma_{\text{prior}}$ are the mean and standard deviation of the prior distribution
- $\mu_{\text{data}}, \sigma_{\text{data}}$ are the mean and standard deviation of the distribution fitted from the data only
- n is the number of data values

2.1.5 Design values

Design values are used when one aims to calculate 'safe' results. This means that outputs, such as a safety factor or the required strength of a construction, should be based on unfavorable conditions – meaning unfavorable input values. In many cases, designs are based on results obtained under these conditions.

Design values in the Probabilistic Library are derived from the stochastic definition as follows:

$$V_{\text{design}} = \frac{Q(q)}{\gamma} \quad (2.30)$$

where:

- V_{design} is the design value
- Q is the function, which calculates the quantile value
- q is the user given design quantile value
- γ is the user given design factor

2.1.6 Goodness of fit

The goodness of fit is an indication how well data fit into the distribution. The Kolmogorov-Smirnov statistic is a goodness of fit test, which is defined as the maximum difference between the CDF values of the derived distribution and the CDF values of the data. The Kolmogorov-Smirnov statistic is calculated as follows:

$$k = \max_{i,\delta \in \{0,1\}} | F(x_i) - \frac{i - \delta}{N} | \quad (2.31)$$

The p -value is calculated as follows for high values of N :

$$\lim_{N \rightarrow \infty} p = 1 - \frac{k}{\sqrt{N}} \quad (2.32)$$

where:

i	is the index number of the data value ($1 \leq i \leq N$)
N	is the number of data values
x_i	is the i^{th} sorted data value
$F(x_i)$	is the CDF value corresponding with x_i

The Kolmogorov-Smirnov statistic is applied to observation sets. This is an indication whether two series of observations originate from the same measurements.

2.2 Correlations

This section describes how correlations are handled in the Probabilistic Library. A theoretical background on the available correlation models can be found in [Appendix B](#).

2.2.1 Correlation factor

The correlation factor ρ_{pq} between two random variables, p and q , indicates the degree to which they are related. The correlation can range from -1 (fully anti-correlated) to 0 (uncorrelated) to 1 (fully correlated). Values between these extremes represent partial correlation.

The Pearson correlation factor, which is used in Probabilistic Library, is calculated as follows:

$$\rho_{p,q} = \frac{\sum_i u_{p,i} u_{q,i}}{\sum_i u_{p,i} \sum_i u_{q,i}} \quad (2.33)$$

where:

i	is the index number of the observed value
p, q	indicate random variables
$u_{p,i}$	is the u -value corresponding with observed value x_i (see Equation (2.18) , note that the distribution of variable p is already derived)

2.2.1.1 Distance based correlation factor

If two components with the same variables have a location, the correlation between these variables can be determined based on their residual correlation and their distance, rather than using [Equation \(2.33\)](#). The correlation factor is then calculated as follows:

$$\rho_{t,p,q} = \rho_{t, \text{rest}} + (1 - \rho_{t, \text{rest}}) \cdot \exp\left(-\frac{D_{p,q}^2}{d_t^2}\right) \quad (2.34)$$

where:

t	indicates a variable type
p, q	indicate components

$\rho_{t, \text{rest}}$	is the user provided rest correlation for a certain variable type t
$D_{p, q}$	is the distance between components p and q
d_t	is the correlation length of variable type t

2.2.2 Processing of correlation factors

Several probabilistic techniques are based on uncorrelated u -values. To account for the effect of correlations, model evaluations should not use the x -values directly converted from u , but rather a transformed counterpart of u that incorporates the correlations. In other words, before a model evaluation is carried out, $u_{\text{uncorrelated}}$ is converted to $u_{\text{correlated}}$. This is done as follows:

$$u_{\text{correlated}} = L(u_{\text{uncorrelated}}) \quad (2.35)$$

where L denotes the lower triangular matrix obtained from the Cholesky decomposition of the correlation matrix $[\rho] = LL^\top$.

2.2.3 Archimedean Copulas

Besides the Gaussian correlation, we also support a few Archimedean copulas: the Clayton, Frank and Gumbel copulas. These copulas can only be defined in sets of two correlated stochastic variables. And it is allowed to have e.g. two variables defined with a Clayton copula and two others with a Frank, Gumbel or Gaussian copula.

The definition of the Frank copula is:

$$C(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \quad (2.36)$$

The definition of the Clayton copula is:

$$C(u, v) = (\max [u^{-\theta} + v^{-\theta} - 1, 0])^{-\frac{1}{\theta}} \quad (2.37)$$

The definition of the Gumbel copula is:

$$C(u, v) = \exp \left(- \left((-\ln u)^\theta + (-\ln v)^\theta \right)^{\frac{1}{\theta}} \right) \quad (2.38)$$

As given above, the copulas Clayton, Frank and Gumbel depend on a parameter θ . For all copulas we have the condition : $\theta <> 0$ and also $\theta \geq 1$ for Gumbel and $\theta \geq -1$ for Clayton.

The copula C is defined for probabilities, where our methods work with standard normal variables. We have u as the independent variable, and we look for v as the dependent variable. The technique to find v for a given u differs for each copula.

For the Frank copula, we can solve the equation $C(u, v) = w$ for v analytically:

$$v = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta} - 1)(e^{-\theta w} - 1)}{e^{-\theta u} - 1} \right) \quad (2.39)$$

For the Clayton copula, we can also solve the equation $C(u, v) = w$ for v analytically:

$$v = \left((w^{-\theta} - u^{-\theta} + 1) \right)^{-\frac{1}{\theta}} \quad (2.40)$$

For very high values ($\theta > 10^8$), the next formula is used to avoid numerical problems:

$$v = u * (1 - (\log(1 - u) + \log(u)) / \text{theta}); \quad (2.41)$$

For the Gumbel copula, no analytical solution is available. Therefore, we use a numerical bisection method to find v such that $C(u, v) = w$.

For the Frank copula, if $\theta > 700$, numerical problems occur due to overflow of the exponential function. In that case, we approximate the Frank copula with the comonotonic copula, which means that $u = v$. For the Gumbel copula, we do the same, but now for $\theta > 400$.

In case of the Gaussian copula, we can use the Cholesky decomposition as explained in the previous section.

2.3 Conversion between standard normal variables (U -domain) and real variables (X -domain)

In practice, it is often advantageous to carry out probabilistic analyses in a standardized space, in which each of the variables are independent. This independence can help simplify the probabilistic techniques. The dependence between variables is reintroduced when the standardized variables are transformed back to the real variables. In Probabilistic Library the standardized space that is used is the standard normal space. This means each of the transformed random variables are normally distributed with mean 0, and standard deviation 1. Each of the random variables in X can be transformed to independent standard normal variables $U = (U_1, \dots, U_n)$. If all the X -variables are mutually independent, this transformation can be described as follows:

$$\Phi(u_i) = F_i(x_i) \Rightarrow u_i = \Phi^{-1}[F_i(x_i)] \Rightarrow x_i = F_i^{-1}(\Phi(u_i)) \quad (2.42)$$

where:

Φ is the standard normal distribution function

F_i is the distribution function (cdf) of X_i , the i^{th} variable in X

Φ^{-1} is the inverse standard normal distribution function

F_i^{-1} is the inverse distribution function of X_i , the i^{th} variable in X

x_i is the realisation of the i^{th} variable in X

u_i is the realisation of the i^{th} variable in U corresponding to the i^{th} variable in X

In this simple case (mutually independent X -variables) the transformation is done such that the non-exceedance probabilities of x_i and u_i are equal:

$$P(U_i \leq u_i) = P(X_i \leq x_i) \quad (2.43)$$

The concept of this transformation is schematically depicted in Figure 2.9.

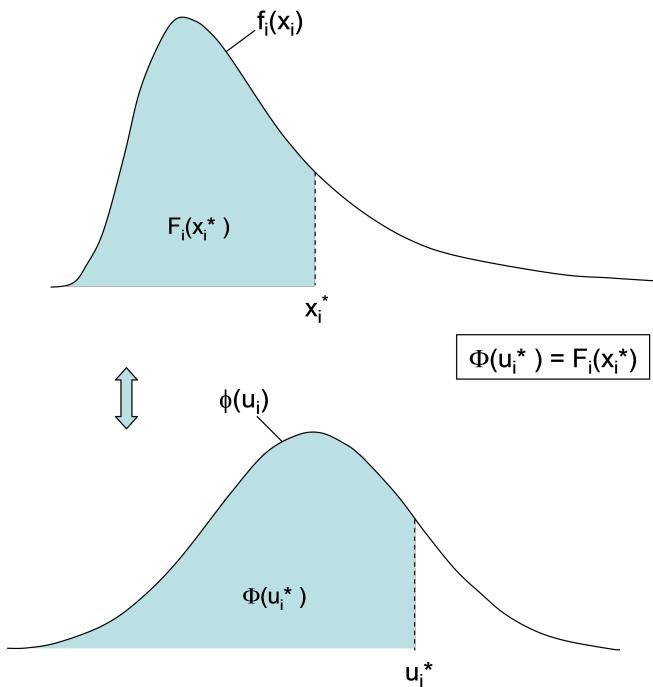


Figure 2.9: Schematic view of the transformation between standard normal variable u_i and real world variable X by means of equal probability of (non-)exceedance.

If the X -variables are not mutually independent, the transformation becomes more complex. In that case the transformation is done with the following conditional probability functions:

$$\begin{aligned}\Phi(u_1) &= F_1(x_1) \\ \Phi(u_2) &= F_2(x_2|x_1) \\ &\vdots \\ &\vdots \\ \Phi(u_n) &= F_n(x_n|x_1, \dots, x_{n-1})\end{aligned}\tag{2.44}$$

This transformation is referred to as the Rosenblatt transformation (cf. Rosenblatt (1952)). The conditional probabilities need to be derived from:

$$f_i(x_i|x_1, \dots, x_{i-1}) = \frac{f_{X_1, \dots, X_i}(x_1, \dots, x_i)}{f_{X_1, \dots, X_{i-1}}(x_1, \dots, x_{i-1})}\tag{2.45}$$

In which $f_{X_i}(x_1, \dots, x_i)$ is a probability density function that is obtained from:

$$f_{X_i}(x_1, \dots, x_i) = \int \dots \int f_X(x_1, \dots, x_n) dx_{i+1} \dots dx_n\tag{2.46}$$

The combination of equations (2.45) and (2.46) potentially requires a cumbersome (numerical) integration procedure.



Note: in practice, this process is far less complex as correlation is usually limited to pairs of variables. Generally, the pairs are split into a dependent variable and an independent variable, and the dependent variable is written as a function of the independent variable. The function consists of a fully deterministic dependent part and a probabilistic independent part:

$$F_{X_2}(x_2|x_1) = G(x_1) + F_2(x_2^*) \quad (2.47)$$

where G is a deterministic function, F_2 a probability distribution function and x_2^* a newly introduced random variable that represents the part of variable x_2 that is independent of x_1 . This type of correlation models is schematically depicted in Figure 2.10. For now, it is important to note that with a correlation model as described by the general form of equation (2.47), the transformation from and to the U -space for variable X_2 is done for the independent part:

$$\Phi(u_2) = F_2(x_2^*) \quad (2.48)$$

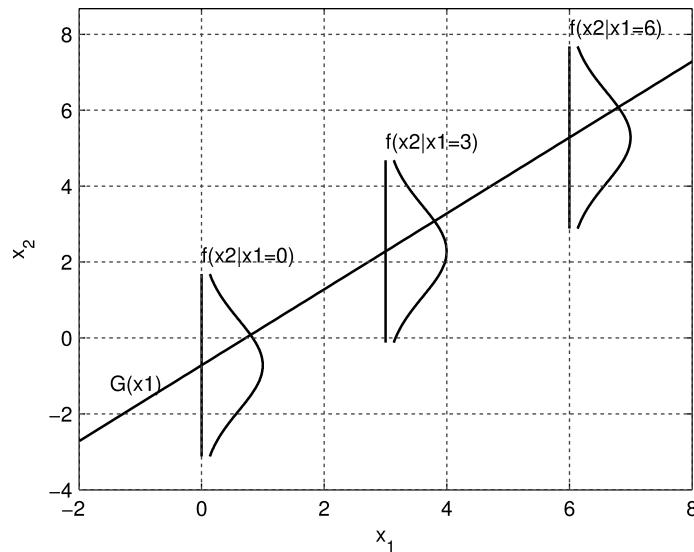


Figure 2.10: Schematic view of correlated variables x_1 and x_2



Note: not all random variables can be represented in a meaningful way by U -variables. This is for instance the case for cyclic variables like wind direction. The reason is that transformation (2.42) only makes sense if large (extreme) u -values are associated with large (extreme) x -values. Potential outcomes of cyclic variables are not ordered from small to large. The domain for these variables is 1 – 360 degrees, but 360 degrees is not larger than 1 degree. Actually, 1 degree is almost the same as 360 degrees. Deltares (2023) describes how such cyclic variables are treated.

3 Sensitivity

Sensitivity analysis is a key tool for understanding how variations in input parameters affect the output of a model. By quantifying the influence of different variables, it helps identify the most critical factors, improve model robustness, and support decision-making under uncertainty.

This chapter presents the methods available in the Probabilistic Library for conducting a sensitivity analysis.

3.1 Single value variation

In single-value variation, variables are varied one at a time and set to a high and low value, while non-varied variables are fixed at their median ($u = 0$).

The main drawback of this method is that it does not account for the combined effects of input variables.

3.2 Sobol indices

Sobol indices quantify the influence of a parameter using a Monte Carlo simulation [Saltelli *et al.* (2010)]. By mixing sample values from an initial Monte Carlo run, the Sobol indices are determined as follows.

The first step consists of performing two Monte Carlo simulations, each with N samples, resulting in sets A and B . From the combined set, the variance v of the model results is calculated.

Next, for each variable p , new sample sets A^p and B^p are generated, with the value for sample i given by:

$$A^p = u_{i,j,A} = \begin{cases} j = p & u_{i,j,B} \\ j \neq p & u_{i,j,A} \end{cases} \quad (3.1)$$

and

$$B^p = u_{i,j,B} = \begin{cases} j = p & u_{i,j,A} \\ j \neq p & u_{i,j,B} \end{cases} \quad (3.2)$$

Then, the model values z corresponding to the samples u are calculated. The first-order index and total index for variable p are computed as follows:

$$I_{\text{first order}} = \frac{1}{N} \cdot \frac{1}{v} \cdot \sum_{i=1}^N z(A_i) z(B_i^p) - z(A_i) z(B_i) \quad (3.3)$$

and

$$I_{\text{total}} = \frac{1}{2N} \cdot \frac{1}{v} \cdot \sum_{i=1}^N [z(A_i) - z(A_i^p)]^2 \quad (3.4)$$

where:

- i is the index number of the sample
- j is the variable index
- A_i is the i^{th} sample in sample set A
- $z(A_i)$ is the model result of the i^{th} sample in sample set A
- N is the number of samples in a Monte Carlo simulation
- v is the variance of the model results of the combined set A and B

The total index I_{total} represents the overall influence of the parameter on the model result.

4 Output uncertainty

Output uncertainty is an extension of sensitivity analysis. In this approach, all input parameters are varied according to their defined uncertainties, leading to uncertainty in the output parameters.

This method is useful when one is interested in the potential future values of a physical property. For example, due to a load, subsidence of the soil surface may occur. It would be useful to know how much subsidence could take place over the next ten years. Another example is the movement of a gully due to a side stream. In this case, it is important to predict the gully's location in the coming year.

This chapter presents the methods available in the Probabilistic Library for output uncertainty analysis.

4.1 Numerical Integration

Numerical Integration (see also [section 5.3.1](#)) varies all input parameters step by step, generating and calculating a realization for each step combination. The resulting realizations are collected into bins, forming a histogram distribution.

4.2 Crude Monte Carlo

Crude Monte Carlo (see also [section 5.3.3](#)) is the 'standard' Monte Carlo simulation method. Random realizations are generated, and their results are categorized into bins. The more realizations a bin contains, the higher its estimated probability.

The Crude Monte Carlo simulation produces a histogram distribution (see [section 2.1.3.25](#)) or, if all model results are identical, a deterministic distribution (see [section 2.1.3.1](#)). The distribution is fitted using all model results.

4.3 Importance Sampling

Importance Sampling modifies randomly selected realizations so that more realizations are concentrated in a user-specified region. Each realization is transformed into another realization using the same approach as Importance Sampling for reliability (see [section 5.3.4](#)).

This method is particularly useful when one is interested in the tail of the distribution.

4.4 Directional Sampling

Directional Sampling (see also [section 5.3.7](#)) is an accelerated sampling method that identifies the failure domain by sampling along randomly chosen directions. The method operates in a transformed standard space with standard Gaussian variables (mean = 0, standard deviation = 1).

After transferring the model into the standard space, the value of model Z_0 is calculated for a sample at the center of space u_0 .

n points are randomly drawn in the variable space following a uniform distribution. The lines connecting these points to the origin serve as the random directions. In the next step, the required reliability index β_q corresponding to the given quantile value or probability of exceedance, is evaluated and used to determine the distance $d_{i,0} = |\beta_q|$ from the origin. The distance is then used to obtain n sample points in the sample space.

Now, the value of the model $Z_{i,1}$ at each sample point ($d_{i,0}$ distance from the origin on each direction) is evaluated. Subsequently, the related β value for each samples (i.e. directions) is calculated. If β_q is greater than β_0 , every direction with $Z_{i,1} < Z_0$ is disregarded in the future probability calculations (i.e., they proceed in an undesired direction), otherwise, they are considered valid. Note that the number of samples n is independent of their validity status. To calculate the probability of failure, equation [Equation \(5.31\)](#) will be used. At this stage, the assumption of $\beta_{dir} = d_{i,0}$ for all directions is applied. See [4.1](#) as the first step for generating the directions and samples, shown for the model $Z = X_1 X_2$, with X_1 and $X_2 = N(0, 1)$.

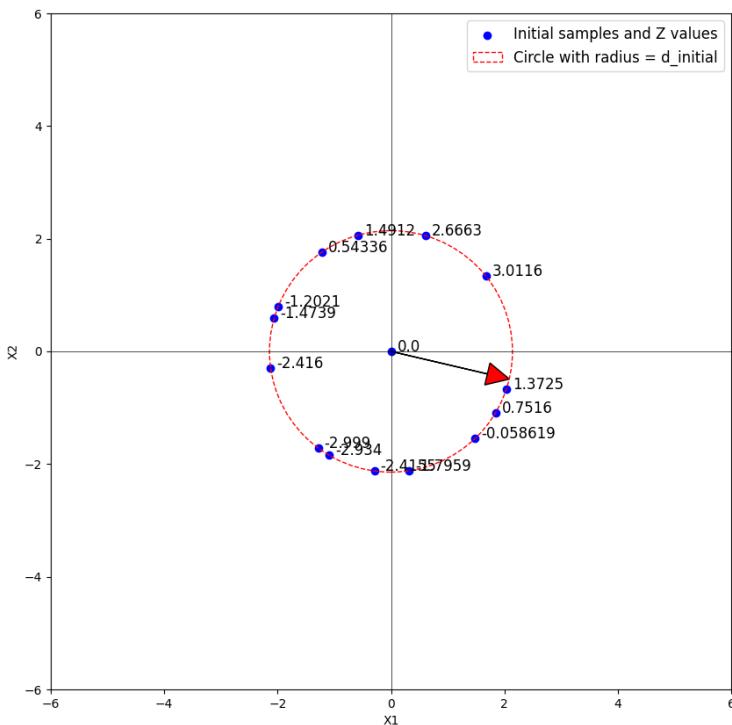


Figure 4.1: Generation of first samples in random directions

The main goal is to find the same value of Z in all directions, with a given tolerance, as the model response for a given exceedance probability. If the Z are not the same for all directions, the following steps must be followed to achieve consistency.

[1] - For each direction i , determine a new distance $d_{i,2}$ from the origin such that it satisfies the following condition:

- ◊ the predicted model value Z_{pred} is the same for all directions (prediction based on interpolation between Z_0 and $\max(Z_{i,1})$) ;
- ◊ the combined reliability, based on equation [Equation \(5.30\)](#), remains equal to β_q . Note that even when a direction is disregarded in previous steps, still the number of directions n , is not changed.

[2] - $d_{i,2}$ is calculated based on the interpolation between Z_{pred} and $Z_{i,1}$ values. Now the model response $Z_{i,2}$ at distance $d_{i,2}$ is evaluated.

The final steps will be repeated to reach Z_i , either up to a maximum of iterations or until the convergence criteria are met. The final Z_i then used as the output of the model at the given quantile value or probability of exceedance. The graph 4.2 shows the output for 90% quantile, where $Z = 1,3826$ is evaluated. The reader may notice how even a limited number of valid samples shape the borders of the failure area for the given exceedance quantile, forming a linear line for the simple model of $Z = X_1 + X_2$.

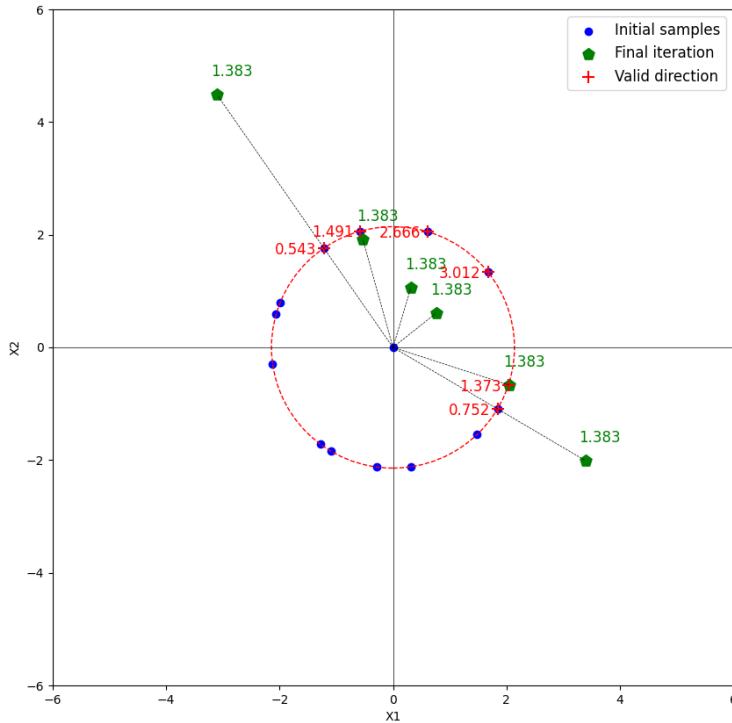


Figure 4.2: The second/final step of iterations for the valid directions

For the error criteria, the distribution of differences between the distances to the center of the U -space in all valid directions from the last and current iterations is considered. These differences form a distribution, and its 95% quantile is used as an indicator of convergence. This quantile is then compared with a user-defined threshold.

4.5 FORM

FORM (First Order Reliability Method) is like FORM in reliability analysis (see also [section 5.3.10](#)). Starting from the origin in u -space, steps of fixed length are taken along the steepest gradient. For each step a realization is performed and its result is added to a CDF curve (see [section 2.1.3.24](#)).

Each subsequent step in the resulting CDF curve is calculated as follows:

$$\vec{u}_{i+1} = \vec{u}_i + L \cdot \nabla z(\vec{u}_i) \quad (4.1)$$

The reliability of the added point is:

$$\beta_{i+1} = \|\vec{u}_{i+1}\| \quad (4.2)$$

where:

\vec{u} is a point in the parameter space defined in u-values

$z(\vec{u})$ is the z value calculated at \vec{u}

L is the step size

$\nabla z(\vec{u})$ is the steepest gradient of z at \vec{u}

4.6 FOSM

FOSM (First Order Second Moment) is a much faster technique than Crude Monte Carlo. It calculates the gradient at the origin and then predicts the shape of the resulting distributions. It assumes that the results follow a normal distribution.

Although very fast, its assumption that results have a normal distribution is often not true. However, for quickly estimating the output uncertainty, it can be very useful.

5 Reliability

Reliability analysis is essential for assessing the performance and safety of engineering systems under uncertainty. By evaluating the likelihood of failure and identifying critical factors influencing system behavior, reliability analysis supports risk-informed decision-making.

This chapter covers the reliability methods available in Probabilistic Library for assessing the reliability of both individual components and systems of multiple components. Additionally, it explores methods for upscaling reliability, extending assessments from small-scale components to a larger scale.

5.1 Probability of failure of a single component

The probability of failure for a single component can be written formally as follows:

$$P_f = P(Z(X_1, X_2, \dots, X_n) < 0) \quad (5.1)$$

where:

P_f is the failure probability

Z is the limit state function

X is the vector of random variables

The limit state function, Z , defines failure in terms of load and strength variables such that $Z < 0$ represents failure. Function Z is often denoted:

$$Z = R - S \quad (5.2)$$

where:

R is the strength, or resistance

S is the load



Note: in case of a flood defense, the load typically consists of the combination of water levels and waves and in some cases currents. The strength is a combination of dike characteristics that reflect the ability of the dike to resist high loads. A simple example is overflow of a river levee in a river (see [Figure 5.1](#)). The load in this simple example is the water level in the river, h_{river} , and the strength of the levee is captured by its height, h_{levee} . The limit state function for this example is simply:

$$Z = R - S = h_{levee} - h_{river} \quad (5.3)$$



Figure 5.1: Illustration of a river levee exposed to the threat of overflow.

The limit state function, Z , is a function of a number of random variables representing both load and strength variables. The probability of failure can be written as follows:

$$P_f = \int \int_{Z<0} \dots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 dx_2 \dots dx_n = \int_{Z<0} f_X(x) dx \quad (5.4)$$

where f is the joint probability distribution of the random variables and X is the vector of variables: $X = (X_1, \dots, X_n)$. Note that random variables are typically denoted with a capital letter (X), while potential realizations of the random variables are denoted with lower case letters (x).

While an analytical solution to equation (5.4) would be ideal, it is typically not available because the Z -function is too complex. Therefore, the probability of failure needs to be estimated with probabilistic computation techniques. Different techniques are available for this purpose. These techniques will be described in detail in the [section 5.3](#).

5.2 Reliability index β

The reliability index, β , is a measure for the reliability of a system, i.e. a measure for the probability of failure of the system. Similar to the probability of failure, β is often defined for a given period of time, e.g. a year. The reliability index is best explained by an example in which the resistance, R , and load, S , of the system are both described as the sum of independent normally distributed random variables. The sum of a set of independent normally distributed random variables is also a random variable (see, e.g. [Grimmett and Stirzaker \(1983\)](#)). This means in this case, R and S are also normally distributed variables and the same can be stated about the Z -function ($Z = R - S$). Define μ_R , μ_S and μ_Z as the respective mean values of R , S and Z and σ_R , σ_S and σ_Z as the respective standard deviations of R , S and Z . The following relations hold:

$$\begin{aligned} \mu_Z &= \mu_R - \mu_S \\ \sigma_Z &= \sqrt{\sigma_R^2 + \sigma_S^2} \end{aligned} \quad (5.5)$$

Figure 5.2 shows an example with $\mu_R = 6$, $\mu_S = 2$, $\sigma_R = \sigma_S = 1$ and consequently $\mu_Z = 4$ and $\sigma_Z = \sqrt{2}$. In reliability analysis, μ_Z is generally a positive value because

otherwise, failure ($Z < 0$) would occur even during "average" conditions. Since Z is normally distributed, $P[Z < 0]$ is equal to:

$$P(Z < 0) = \Phi\left(-\frac{\mu_Z}{\sigma_Z}\right) = 1 - \Phi\left(\frac{\mu_Z}{\sigma_Z}\right) \quad (5.6)$$

where Φ is the standard normal distribution function. Based on this equation, the reliability index, β , is defined as:

$$\beta = \frac{\mu_Z}{\sigma_Z} \quad (5.7)$$

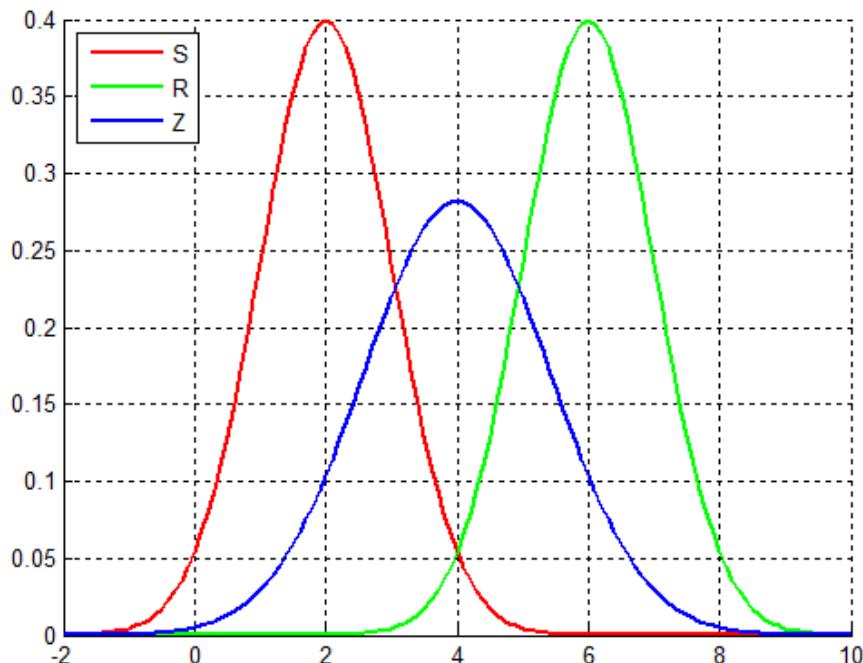


Figure 5.2: Example of normally distributed S , R and Z functions with $\mu_R = 2$, $\mu_S = 6$, $\sigma_R = 1$, $\sigma_S = 1$, $\mu_Z = 4$ and $\sigma_Z = \sqrt{2}$.

For the specific case where the Z -function is normally distributed, the relation between β and the probability of failure is:

$$P(Z < 0) = \Phi(-\beta) = 1 - \Phi(\beta) \quad (5.8)$$

Or inversely:

$$\beta = \Phi^{-1}(1 - P(Z < 0)) \quad (5.9)$$

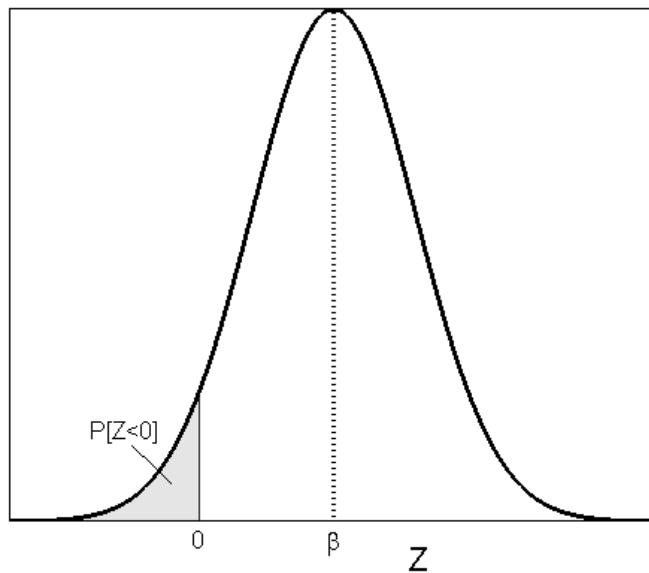


Figure 5.3: Schematic view of the relation between the reliability index β and the probability of failure, $P[Z < 0]$. Z is normally distributed with mean β and standard deviation 1

This shows why the reliability index β is a measure for reliability. Note that if the Z -function is not normally distributed, equation (5.8) does not necessarily hold if β is defined as in equation (5.7). This is why β is a measure for reliability, not an exact representative of the probability of failure. However, often β is computed directly from equation (5.9) in which case it is an exact representative of the probability of failure by definition. Table 5.1 shows a range of β -values and associated probabilities of exceedance.

Table 5.1: Values of reliability index β and associated probability of failure.

β	P
1.0	$1.59 \cdot 10^{-1}$
1.5	$6.68 \cdot 10^{-2}$
2.0	$2.28 \cdot 10^{-2}$
2.5	$6.21 \cdot 10^{-3}$
3.0	$1.35 \cdot 10^{-3}$
3.5	$2.33 \cdot 10^{-4}$
4.0	$3.17 \cdot 10^{-5}$
4.5	$3.40 \cdot 10^{-6}$
5.0	$2.87 \cdot 10^{-7}$

Small values of β indicate large probabilities of failure, large values of β indicate small probabilities of failure. This can be easily explained with some examples in which, for the sake of simplicity, Z is assumed to be normally distributed. If $\beta = 1$, failure occurs if a random sample of Z is more than $1 \cdot \sigma_Z$ lower than the mean. The probability for this to happen is equal to $\Phi(-1) \approx 0.16$. On the other hand, if $\beta = 4$, failure occurs if a random sample of Z is more than $4 \cdot \sigma_Z$ lower than the mean. The probability for this to happen is equal to

$\Phi(-4) \approx 3.2 \cdot 10^{-5}$. In other words: larger values of β indicate that more extreme events are required for failure to occur, hence a lower probability of failure, hence a larger reliability of the system.

5.3 Reliability methods

Equation (5.4) describes the general formulation of the probability of failure of a single component. While an analytical solution to equation (5.4) would be ideal, it is typically not possible because the Z -function is too complex. Therefore, the probability of failure needs to be estimated with probabilistic computational techniques. The computational techniques available within the Probabilistic Library are summarized in this section.

The reason to implement a set of probabilistic computational techniques in the Probabilistic Library is that each technique has its (dis)advantages with respect to criteria such as robustness, accuracy and required computation time. The "best" technique to be applied therefore depends on the problem under consideration.

A theoretical background of the most well-known computation techniques is presented in [Appendix A](#).

5.3.1 Numerical Integration

Numerical integration is the most time-consuming but most accurate method for calculating failure probability. A step size, along with the minimum and maximum values of the input variables, is required for the evaluation.

If the underlying model fails for a certain realization, its results will be ignored. The user must determine whether the number of failed realizations is acceptable.

The minimum and maximum values for which the integration runs are defined in the u -space. Numerical integration fills the remaining space between these values and -8 and 8 with additional cells to ensure complete coverage of the integration domain.

Note: $u = 8$ corresponds to a probability of approximately 6×10^{16} , which is orders of magnitude smaller than the probabilities of typical scenarios.

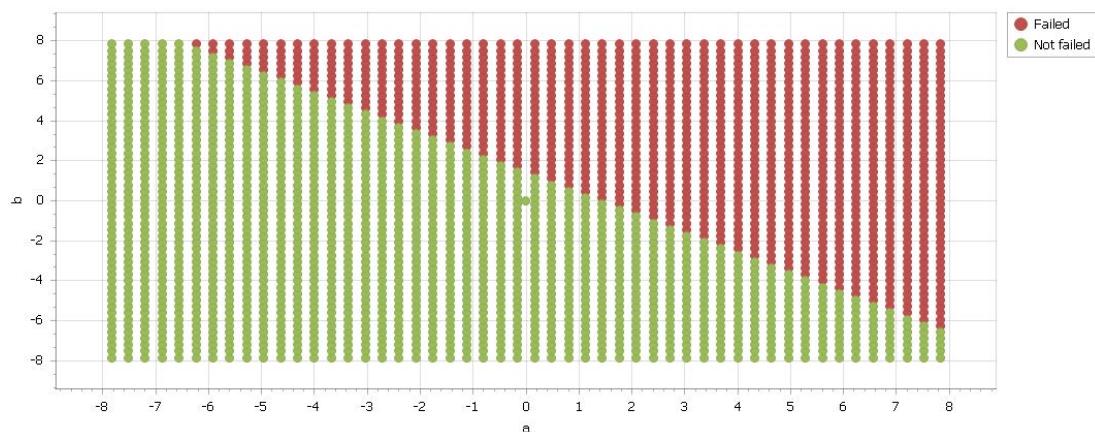


Figure 5.4: Numerical integration: realizations for two stochastics

5.3.2 Numerical Bisection

Numerical Bisection is similar to Numerical Integration, but in this method, the integration cells are generated by bisecting the entire domain.

Whenever an integration cell has the same qualitative result (fail, not fail, or not counting) at all its corner points, we assume that the result remains the same throughout the cell. Consequently, no additional calculations are performed inside the cell or along its edges.

If a cell contains corner points with different qualitative results, it is split into smaller cells, and new corner points are evaluated.

This process continues until the remaining cells – those with mixed results – represent a probability lower than an acceptable threshold in reliability. The user can specify this threshold.

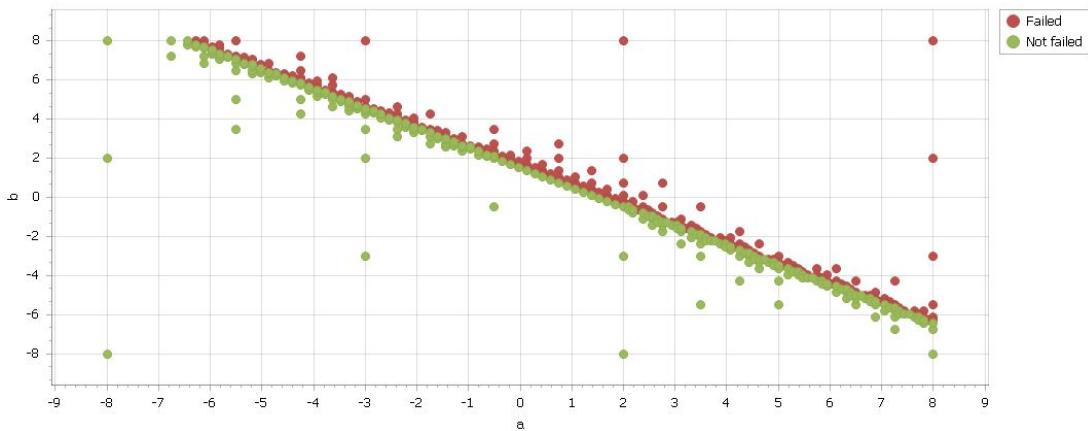


Figure 5.5: Numerical Bisection: realizations for two random variables

5.3.3 Crude Monte Carlo

5.3.3.1 Algorithm

Crude Monte Carlo is the standard Monte Carlo simulation method. Random realizations are generated in proportion to their probability density. The number of realizations that result in failure is counted, as well as the number that do not lead to failure. The probability of failure is then calculated as follows:

$$p_{\text{failure}} = \frac{N_{\text{failure}}}{N} \quad (5.10)$$

where:

- p_{failure} is the probability of failure
- N_{failure} is the number of realizations which were interpreted as failing
- N is the total number of realizations

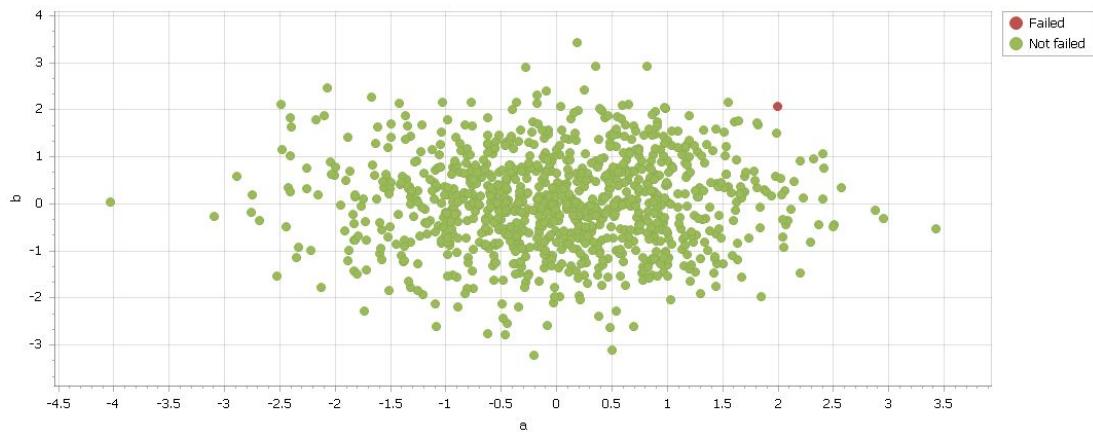


Figure 5.6: Crude Monte Carlo: realizations for two random variables

5.3.3.2 Convergence

The Monte Carlo simulation also leads to a standard deviation of the probability of failure. This standard deviation σ_p is based on a confidence level α :

$$\sigma_p = z \sqrt{\frac{p(1-p)}{N}} \quad (5.11)$$

with:

$$z = \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \quad (5.12)$$

where:

- σ_p is the standard deviation of the probability of failure
- p is the probability of failure p_{failure}
- N is the total number of realizations
- z is the quantile of the standard normal distribution corresponding with α (z is equal to 1 in the Probabilistic Library)
- Φ is the CDF of the standard normal distribution (see [Equation \(2.17\)](#))
- α is the confidence level

The standard deviation σ_p can be expressed as a variation coefficient ε :

$$\varepsilon = \begin{cases} p < \frac{1}{2} & \frac{\sigma_p}{p} = z \sqrt{\frac{1-p}{Np}} \\ p \geq \frac{1}{2} & \frac{\sigma_p}{(1-p)} = z \sqrt{\frac{p}{N(1-p)}} \end{cases} \quad (5.13)$$

The user can set the maximum variation coefficient ε_{\max} . The Monte Carlo simulation will stop when $\varepsilon \leq \varepsilon_{\max}$ (and limited to the minimum and maximum number of realizations). The confidence level α cannot be set by the user. Its value is such that z is 1.

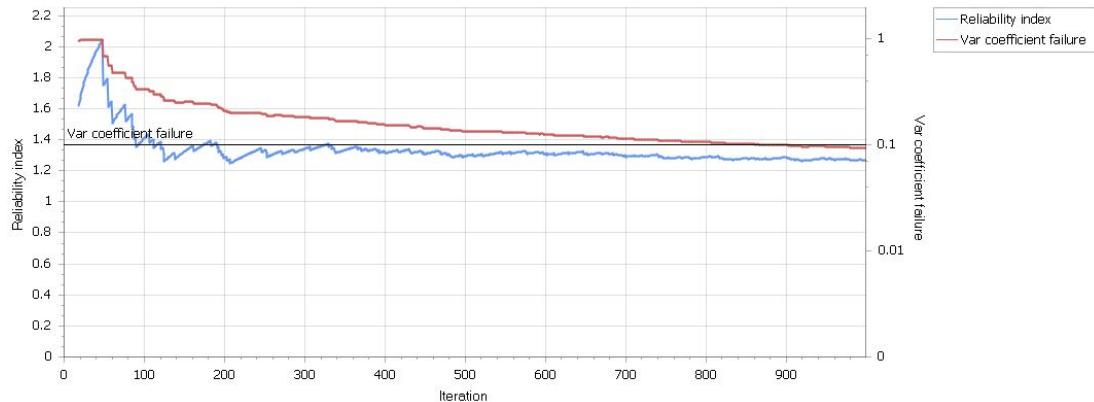


Figure 5.7: Crude Monte Carlo convergence

When the underlying model fails for a certain realization, its results are ignored. The user should decide whether the number of failed realizations is acceptable.

The minimum and maximum values for which the integration will run are defined in the u -space. The Monte Carlo analysis will then check whether the remaining area leads to failure or not.

5.3.4 Importance Sampling

5.3.4.1 Algorithm

Importance sampling is a method to increase the efficiency of the Crude Monte Carlo method; that is, to decrease the number of samples and Z -function evaluations required to produce a reliable estimate of the failure probability. This is done by replacing the initial probability density of the input variables by a more efficient one, in which "efficient" refers to the proportion of the samples which will result in failure. An increasing percentage of samples in the failure domain results in a reduction in the variance of the estimator of the failure probability, hence a smaller number of samples is required for a reliable estimate.

In Crude Monte Carlo, realizations are drawn proportionally to their probability density $\varphi(u_i)$, since the Probabilistic Library uses the standard normal space. With Importance Sampling, each realization u is mapped to u_{imp} . The Probabilistic Library supports the following translation for each variable individually:

$$u_{\text{imp}} = \mu_{\text{var}} + \sigma_{\text{var}} \cdot u_{\text{var}} \quad (5.14)$$

where:

- μ_{var} is the user defined mean value per variable. The combination of all variables is the mean realization of the Importance Sampling algorithm;
- σ_{var} is the user defined standard deviation per variable.

A correction is applied in the calculation of failure to compensate for this translation. This is done by giving each realization a weight, which is calculated as follows (the multiplication with σ_{var} is performed to compensate for the dimensionality):

$$w_{\text{var}} = \frac{\sigma_{\text{var}} \cdot \varphi(u_{\text{imp}})}{\varphi(u_{\text{var}})} \quad (5.15)$$

and

$$W_{\text{realization}} = \prod_{\text{variables}} w_{\text{var}} \quad (5.16)$$

where:

φ is the standard normal probability density function, see [Equation \(2.16\)](#)

w_{var} is the weight factor per variable per realization

$W_{\text{realization}}$ is the weight factor of the realization

Corresponding to [Equation \(5.10\)](#), the probability of failure is equal to:

$$p_{\text{failure}} = \frac{\sum_{\text{failing realizations}} W}{\sum_{\text{all realizations}} W} \quad (5.17)$$

To avoid numerical issues and because:

$$\lim_{N \rightarrow \infty} \sum_{\text{all realizations}} W = N \quad (5.18)$$

The probability of failure is calculated as follows:

$$p_{\text{failure}} = \frac{\sum_{\text{failing realizations}} W}{N} \quad (5.19)$$

5.3.4.2 Convergence

Corresponding to [Equation \(5.11\)](#) and [Equation \(5.13\)](#), the standard deviation and variance coefficient become:

$$\sigma_p = z \sqrt{\frac{p(W_{\text{design point}} - p)}{N}} \quad (5.20)$$

and

$$\varepsilon = \begin{cases} p < \frac{1}{2} & \frac{\sigma_p}{p} = z \sqrt{\frac{W_{\text{design point}} - p}{Np}} \\ p \geq \frac{1}{2} & \frac{\sigma_p}{(1-p)} = z \sqrt{\frac{W_{\text{design point}} - (1-p)}{N(1-p)}} \end{cases} \quad (5.21)$$

where:

z is a value related to the confidence level (see [Equation \(5.12\)](#)). In the Probabilistic Library, z is 1

$W_{\text{design point}}$ is the weight of the realization of the design point

N is the total number of realizations

When the underlying model does not succeed for a certain realization, its results will be ignored. The user should decide whether the number of non succeeded realizations is acceptable.

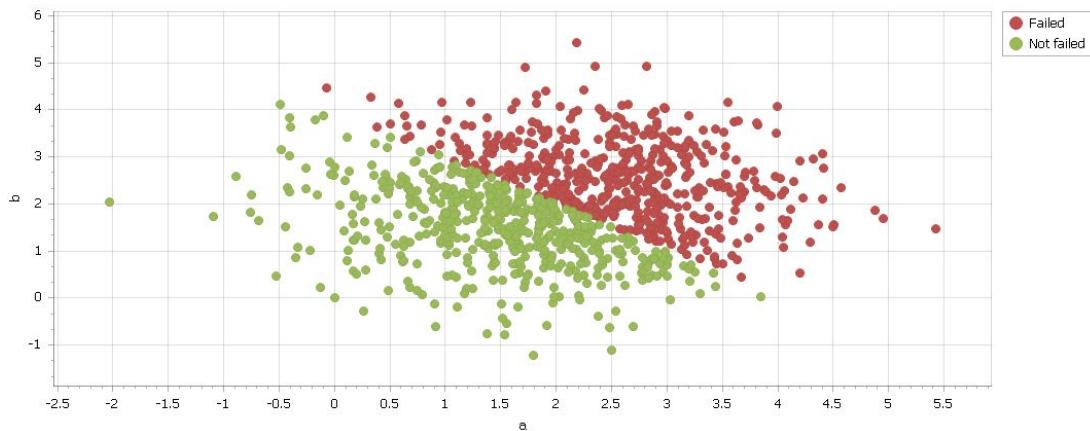


Figure 5.8: Importance Sampling: realizations for two random variables

5.3.4.3 Mean realization

The mean realization (see [Equation \(5.14\)](#)) is essential for the successful behaviour of the Importance Sampling algorithm. The mean realization can be user defined or derived automatically. The following options are available:

None The mean realization is user defined.

Direction Along a user defined direction is searched until the limit state is found, limited by a maximum distance from the origin. This option still requires some knowledge where to find the mean realization.

Sphere Over a sphere is searched until the limit state is found. Only at angles 0° , 90° , 180° and 270° is searched. If the limit state is not found, the radius of the sphere is increased. This is a time consuming option.

Sensitivity Along the direction of the steepest gradient is searched until the limit state is found, limited by a maximum distance from the origin. The steepest gradient is found by taking the gradient for each variable individually and then combining them. This is the preferred option.

5.3.5 Adaptive Importance Sampling

5.3.5.1 Algorithm

Adaptive Importance Sampling is an improvement on the Importance Sampling method. In general, Importance Sampling is sensitive to the user-provided starting point. If the starting point is not chosen carefully, the method may yield unsatisfactory results. In Adaptive Importance Sampling, iterative loops are used to refine the starting point. To achieve this, the K-Means algorithm is applied to identify multiple starting points based on the samples that led to failure in the previous iteration.

The Adaptive Importance Sampling algorithm is displayed in Figure 5.9.

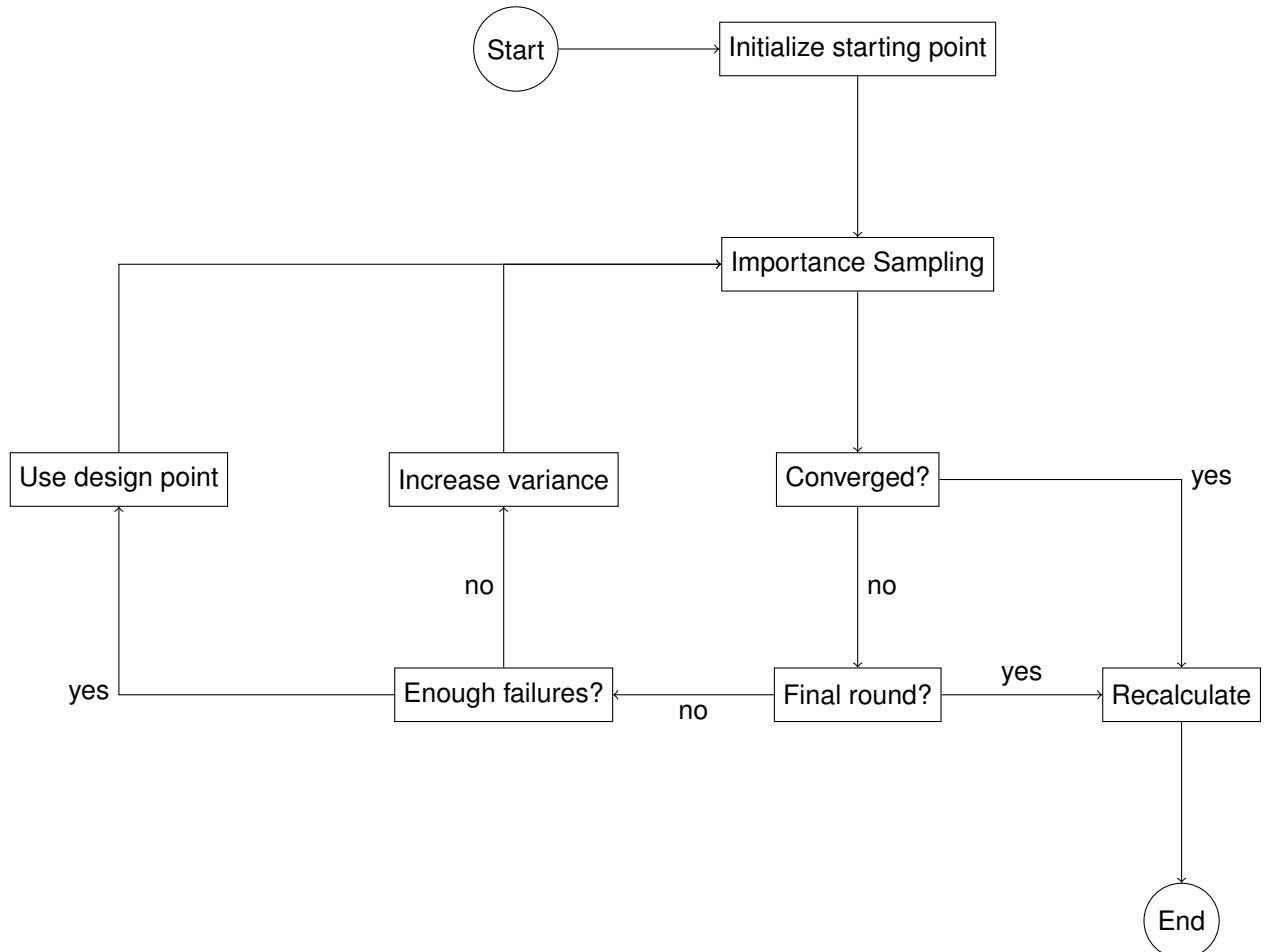


Figure 5.9: Adaptive Importance Sampling

These loops are executed until a required fraction of failed realizations $\varepsilon_{\text{failed}}$ is reached:

$$\min \left(\frac{N_{\text{failed}}}{N}, 1 - \frac{N_{\text{failed}}}{N} \right) \geq \varepsilon_{\text{failed}} \quad (5.22)$$

where:

- N is the total number of failed realizations
- N_{failed} is the number of failed realizations

The following is performed in each step:

Initialize starting point	Find the starting point with one of the starting point algorithms (see section 5.3.4.3).
Importance Sampling	The Importance Sampling algorithm (see section 5.3.4.1). The number of samples in each Importance Sampling can be set to a maximum. When the auto break option is used, the algorithm decides when to break a loop and continue with the next loop (see section 5.3.5.2).
Converged?	Checks whether there is convergence (see Equation (5.22)).
Final round?	Checks whether the last allowed round is reached, in order to prevent endless loops.
Recalculate	Recalculates the last round with a possibly higher number of realizations.
Enough failures?	Checks whether enough failures are found. Depending on this value, the kind of modification of Importance Sampling settings is determined.
Increase variance	Increases the variance (see section 5.3.4.1) for the next loop.
Use design point	Uses the design point found in the last loop as starting point in the next loop. In case no design point was found, the realization closest to the limit state is used.

The following features are available to reduce calculation time:

- ◊ The maximum number of realizations for non-final loops can be specified separately from the maximum number of realizations in the final loop.
- ◊ Subsequent loops can be skipped if the reliability index exceeds a predefined threshold determined in the first loop.
- ◊ When the design point is carried over to the next loop, it will be rounded to a specified value. Specify 0 if no rounding is desired.
- ◊ When using scenario tables, previous calculation results can be reused.

5.3.5.2 Auto break loops

The number of realizations in the loops in Adaptive Importance Sampling can be determined automatically. This algorithm is based on the convergence criterion $\varepsilon_{\text{weight}}$, which specifies the maximum relative weight of all failing samples. Usually this value is 0.1.

The loop is broken when the required number of runs with the current starting point ($n_{\text{additional}}$) is more than the expected number of runs with an improved starting point (n_{expected}):

$$n_{\text{additional}} > n_{\text{expected}} \quad (5.23)$$

with:

$$n_{\text{additional}} = n_{\text{current}} \cdot \left(\frac{W_{\max}}{W_{\text{total}} \cdot \varepsilon_{\text{weight}}} - 1 \right) \quad (5.24)$$

and

$$n_{\text{expected}} = \frac{2(\beta_{\text{current}} + 1)}{\varepsilon_{\text{weight}}} \quad (5.25)$$

where:

n_{current}	the number of samples already made in the current loop
W_{\max}	the maximum weight of a failed sample in the current loop
W_{total}	the total weight of the failed samples in the current loop
β_{current}	the reliability index based on the samples in the current loop
$\varepsilon_{\text{weight}}$	convergence criterion, which specifies the maximum relative weight of all failing samples, default is 0.1

5.3.6 Subset simulation

Subset Simulation [Au (2001)] is a staged version of Crude Monte Carlo (see Section 5.3.3). It consists of multiple iterations, with the first iteration identical to Crude Monte Carlo. The simulation stops once the Crude Monte Carlo convergence criterion is met, i.e., when $\varepsilon \leq \varepsilon_{\max}$ (see Equation (5.13)).

If the convergence criterion is not met, a new Crude Monte Carlo iteration is performed, using a fraction k of the realizations from the previous run. The realizations closest to failure are selected, and new realizations are generated from them using a Markov Chain or Adaptive Conditional Sampling [Papaioannou *et al.* (2015)].

For each old realization, $\frac{1}{k}$ new realizations are generated. A new realization is generated in the following way, with u -values for each variable var:

$$u_{\text{var, new}} = \begin{cases} r_{\text{var}} \geq R_{[0, 1]} & u_{\text{var, prop}} \\ r_{\text{var}} < R_{[0, 1]} & u_{\text{var, prev}} \end{cases} \quad (5.26)$$

with:

$$r_{\text{var}} = \frac{\varphi(u_{\text{var, prop}})}{\varphi(u_{\text{var, prev}})} \quad (5.27)$$

and

$$u_{\text{var, prop}} = u_{\text{var, prev}} + R_{[-1, 1]} \cdot \Delta\sigma \quad (5.28)$$

where:

$u_{\text{var, prev}}$	is the u -value per variable of the previous realization
$u_{\text{var, prop}}$	is a proposed new u -value per variable
$u_{\text{var, new}}$	is the new u -value per variable
$R_{[0, 1]}$	is a random value between 0 and 1
$R_{[-1, 1]}$	is a random value between -1 and 1
r_{var}	is the ratio of probability density between $u_{\text{var, prop}}$ and $u_{\text{var, prev}}$
φ	is the probability density function of a standard normal distribution (see Equation (2.16))
$\Delta\sigma$	is the user given maximum deviation from the previous u -value. In case the option 'Adaptive Conditional' is used, this value is derived automatically and updated during the process

The proposed sample u_{prop} is used if the corresponding z -value is less than the z_k , which is the highest z -value in the subset used to generate new realizations, otherwise the original sample is used.

The probability of failure is calculated as follows:

$$p = k^i \cdot p_{i, \text{MC}} \quad (5.29)$$

where:

i	is the iteration index, the first being 0
k	is the fraction taken from the realizations (recommended value 0.1)
$p_{i, \text{MC}}$	is the probability of failure calculated with Crude Monte Carlo using the realizations of the i^{th} iteration

5.3.7 Directional Sampling

5.3.7.1 Algorithm

Directional Sampling is a type of Monte Carlo method which aims to (strongly) reduce the number of samples in comparison with the Crude Monte Carlo method. In Directional Sampling, realizations are represented as directions, rather than points in the parameter space. Along this direction/vector, the point of failure is identified, and the corresponding distance, β , to the origin is determined. The remaining probability of failure beyond this point is calculated and added to the total probability of failure as follows:

$$p_{\text{failure}} = \frac{\sum w_{\text{dir}}}{N_{\text{realizations}}} \quad (5.30)$$

with:

$$w_{\text{dir}} = \Gamma \left(\frac{N_{\text{variables}}}{2}, \frac{\beta_{\text{dir}}^2}{2} \right) \quad (5.31)$$

where:

- w_{dir} is the weight of a direction or the failure probability in this direction
- β_{dir} is the distance along a direction where the first point of failure is found
- Γ is the upper incomplete regularized gamma function

5.3.7.2 Convergence

The standard deviation of the failing probability is calculated as follows:

$$\sigma_p = \sqrt{\frac{\sum (w_{\text{dir}} - p)^2}{N \cdot (N - 1)}} \quad (5.32)$$

The variation coefficient is calculated as follows:

$$\varepsilon = \begin{cases} p < \frac{1}{2} & \frac{\sigma_p}{p} \\ p \geq \frac{1}{2} & \frac{\sigma_p}{(1-p)} \end{cases} \quad (5.33)$$

The Directional Sampling simulation stops if the variation coefficient ε is less than the user given maximum variation coefficient ε_{max} . Also, the number of minimum and maximum directions should be satisfied.

The minimum and maximum number of iterations refer to an internal procedure to determine the point along the direction where failure occurs.

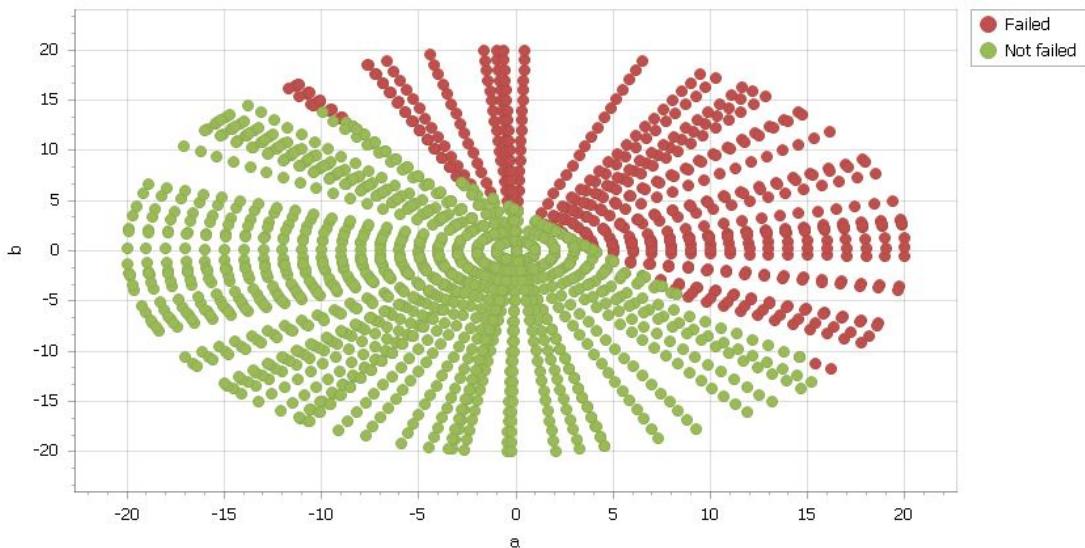


Figure 5.10: Directional Sampling: realizations for two random variables

5.3.8 Latin Hypercube

The Latin Hypercube algorithm is a type of Monte Carlo sampling method. It uses a fixed number of realizations, N . The algorithm divides the u -space into N equal sections, starting from u_{\min} and ending at u_{\max} . It ensures that each section is represented in the realizations, guaranteeing that extreme values are included. However, note that the weights of the realizations may vary. If u_{\min} and u_{\max} differ from -8 and 8 – beyond which calculations are numerically infeasible – the remaining space is treated as a single cell and included in the failure analysis.

Latin Hypercube uses the same convergence definition as Crude Monte Carlo (see Section 5.3.3.2). However, since a fixed number of realizations is used, convergence is calculated only for informational purposes. The main advantage of Latin Hypercube is its computational efficiency. However, it provides only a rough approximation of the probability of failure.

5.3.9 Cobyla

The Cobyla algorithm searches for the point in the u -space that has the highest probability density and indicates failure. This point is assumed to be representative of the design point.

The Cobyla algorithm is an implementation of Powell's nonlinear derivative-free constrained optimization method, which uses a linear approximation approach Powell (1994). It is a sequential trust-region algorithm that employs linear approximations to both the objective and constraint functions. These approximations are formed by linear interpolation at $n + 1$ points in the variable space, and the algorithm aims to maintain a well-shaped simplex throughout iterations.

Because the Cobyla algorithm provides a rough estimation of the reliability and design point, it should be used primarily as an initial method in a reliability study.

5.3.10 FORM

5.3.10.1 Algorithm

The FORM procedure starts from a certain user defined starting point in the parameter space. From there it tries to find a point which is closer to the design point by taking the gradient of the z -value in the u -space of the parameters. The z -value is an indication whether failure occurs and is derived from the failure definition. After a number of steps, the point is close enough to the design point and the calculation stops.

The FORM analysis searches for the design point: when the design point is found, the reliability index β can be evaluated as the distance between the origin and the design point (see Figure 5.11). The corresponding probability of failure is:

$$p_{\text{failure}} = 1 - \Phi(\beta) \quad (5.34)$$

where Φ is the cumulative density function in the standard normal space (see Equation (2.17)).

The probability of failure found in this way is regarded to be a good approximation of the "real" probability of failure.

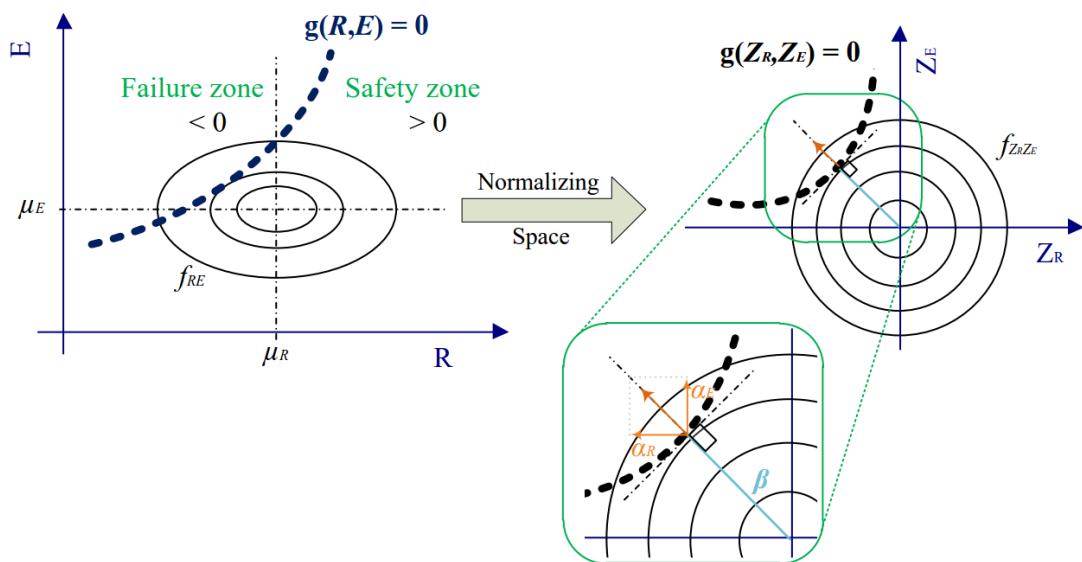


Figure 5.11: Schematic representation of FORM

To find the design point, the FORM analysis starts at a given starting point, usually the origin, in the standard normal space and iterates to the design point.

In each iteration step, the current point is moved closer to the design point. Using the steepest descend, the next point is found, until a convergence criterion is fulfilled. To get from the current point u_i to the next point u_{i+1} , the predicted point u_{pred} is determined.

To calculate u_{pred} , we assume that z is linear with u close to the design point. The gradient is taken between u_i and u_{pred} :

$$\left| \frac{\partial z}{\partial u} \right| = \frac{z_{\text{pred}} - z_i}{u_{\text{pred}} - u_i} = -\frac{z_i}{u_{\text{pred}} - u_i} \quad (5.35)$$

which is equivalent to:

$$u_{\text{pred}} = u_i - \frac{z_i}{\left| \frac{\partial z}{\partial u} \right|} \quad (5.36)$$

where:

- i is the iteration index number;
- $\left| \frac{\partial z}{\partial u} \right|$ is the calculated vector length of the derivative to all variables
- z_{pred} is the limit state value at the design point, which is by definition equal to zero
- z_i is the calculated limit state value at iteration i
- u_{pred} is the predicted vector of the design point
- u_i is the iteration vector of u at iteration i

The step from the current point to the next point is a step into the direction of u_{pred} . The step is not taken completely, but partially with a relaxation factor f_{relax} , in order to prevent numerical instabilities.

$$u_{j, i+1} = -\alpha_{j, i} \cdot |u_{i, \text{pred}}| \cdot f_{\text{relax}} + u_{j, i} \cdot (1 - f_{\text{relax}}) \quad (5.37)$$

with the α value per variable:

$$\alpha_j = \frac{\frac{\partial z_j}{\partial u_j}}{\left| \frac{\partial z}{\partial u} \right|} \quad (5.38)$$

where:

- j is the index of a variable
- $|u|$ is the vector length of the current point
- f_{relax} is the given relaxation factor. The relaxation factor prevents the algorithm from too big steps, which may cause divergence

and finally leads to the reliability index β as:

$$\beta = |u_i| \quad (5.39)$$

The benefit of this method is that it is quick. The disadvantages include the possibility of finding a local design point, which may correspond to a non-representative failure probability. Another disadvantage is that numerical problems may occur since the z -value must be continuous.

It is not always possible to determine whether a design point is a local design point, so the user must apply this method with care and evaluate whether it is suitable for the type of models they are using. Numerical problems can be detected in the convergence chart. If this occurs, one can decrease the relaxation factor or modify the gradient step size (both adjustments may either improve or worsen the results).

5.3.10.2 Convergence

Convergence is reached when the u_i , the value of the last iteration i , is close enough to the predicted value u_{pred} , so:

$$|u_i - u_{\text{pred}}| < \varepsilon \quad (5.40)$$

where:

- u_i is the point at iteration i
- u_{pred} is the predicted final value of u , see [Equation \(5.35\)](#)
- ε is a user given value maximum allowed difference in reliability

and, together with [Equation \(5.40\)](#), the convergence criterion:

$$\frac{|z|}{\left|\frac{\partial z}{\partial u}\right|} < \varepsilon \quad (5.41)$$

Alternatively to [Equation \(5.35\)](#), the gradient is taken between u_0 (all u -values equal to 0) and u_{pred} . Then we get for the linear relation:

$$\left|\frac{\partial z}{\partial u}\right| = \frac{z_{\text{pred}} - z_0}{u_{\text{pred}} - u_0} = -\frac{z_0}{u_{\text{pred}}} \quad (5.42)$$

The value z_0 is calculated with the value for z in the last iteration z_i and the gradient:

$$z_0 = z_i - \sum_j \frac{\partial z_i}{\partial u_j} \cdot u_j \quad (5.43)$$

which leads to:

$$|u_{\text{pred}}| = \frac{|z_0|}{\left| \frac{\partial z}{\partial u} \right|} \quad (5.44)$$

The convergence criterion is defined as the relative difference in length of u_{pred} and u :

$$\frac{|u_{\text{pred}}^2 - u^2|}{u^2} < \varepsilon \quad (5.45)$$

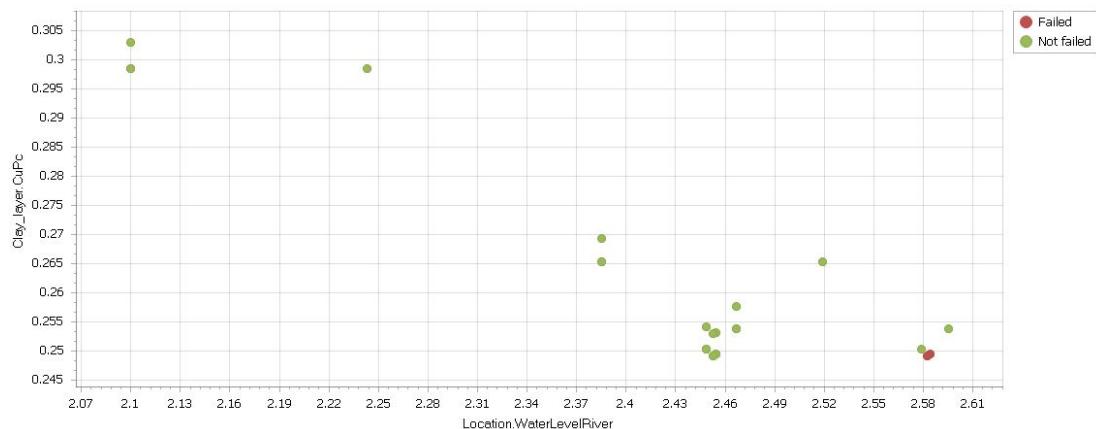


Figure 5.12: FORM: example of iteration path to design point

5.3.10.3 Starting point

The FORM algorithm starts at a specified starting point. In most cases, beginning at the default location is sufficient.

The available starting point options are the same as those for Importance Sampling (see Section 5.3.4.3).

5.3.10.4 Loops

In case calculation options have been specified, which do not lead to convergence, loops can be used to modify the calculation options. If the number of loops is greater than 1, the relaxation factor f_{relax} is modified until convergence is reached. The relaxation factor is modified as follows:

$$f_{\text{relax}} = \frac{1}{2} \cdot f_{\text{relax, prev}} \quad (5.46)$$

where:

$f_{\text{relax, prev}}$ is the relaxation factor in the previous loop.

5.3.11 FORM then Directional Sampling

The "FORM then Directional Sampling" method is a reliability approach based on FORM (see [Section 5.3.10](#)) and Directional Sampling (see [Section 5.3.7](#)).

The method begins with a FORM calculation. If it succeeds (FORM converges), the process is complete, and the FORM result is returned. If FORM fails to converge, a Directional Sampling calculation is performed instead, and its result is used.

The rationale behind this approach is to leverage FORM whenever possible due to its high efficiency while providing a fallback to Directional Sampling in cases where FORM does not converge.

5.3.12 Directional Sampling then FORM

The "Directional Sampling then FORM" method is a reliability approach based on Directional Sampling (see [Section 5.3.7](#)) and FORM (see [Section 5.3.10](#)).

The method begins with a Directional Sampling reliability calculation, followed by a FORM calculation. It returns the reliability index (β) from Directional Sampling and the α -values from FORM. Since FORM starts from the design point identified by Directional Sampling, it typically requires only a few iterations.

The rationale behind this approach is that Directional Sampling quickly estimates the β value but requires many samples to accurately determine all α -values. In contrast, FORM efficiently refines the α -values when initialized close to the design point.

Note that the starting point of FORM depends on the method for "Contribution per variable" (see [section 5.4](#)).

5.4 Contribution per variable

For all numerical integration-based methods (Numerical Integration and Numerical Bisection) and Monte Carlo-based methods (Crude Monte Carlo, Importance Sampling, Directional Sampling, and Subset Simulation), the contribution per variable is calculated after determining the reliability index (β). The applied methods are explained in the following sections.

5.4.1 Center of gravity

The "Center of gravity" method calculates the weighted mean of all failing realizations. This is done as follows:

$$\alpha_j = f_{\text{normal}} \cdot \frac{\sum_{\text{failing realizations } i} u_{i,j} \cdot W_i}{\sum_{\text{failing realizations } i} W_i} \quad (5.47)$$

where:

- i indicates a realization in the applied technique
- j indicates a variable
- u is a realization in the applied technique

W_i is the weight of the realization (for Crude Monte Carlo: 1)
 f_{normal} is a normalizing factor so that $\sum \alpha_j^2 = 1$

5.4.2 Center of angles

In the "Center of angles" method, all realizations are defined using spherical coordinates instead of Cartesian coordinates. The weighted mean of the angles in the spherical coordinates is then calculated and converted back into Cartesian coordinates.

5.4.3 Nearest to mean

The "Nearest to mean" method selects the realization with the highest probability density from all realizations that are considered failing according to the failure definition.

Although this is the fastest method, it is not recommended for Monte Carlo techniques. The results are affected by the inherent randomness of Monte Carlo methods, which can lead to unreliable outcomes. For instance, unimportant variables may be assigned significant α values due to this randomness.

5.5 System analysis

The probabilities of failure of the components are combined to derive the probability of failure of the whole system. This is being referred to as system analysis. System analysis generally deals with parallel systems, series systems or combinations of both. A parallel system refers to a system in which failure only occurs if all components fail. A series system refers to a system where failure occurs if at least one of the components fails. This concept is schematically depicted in [Figure 5.13](#).

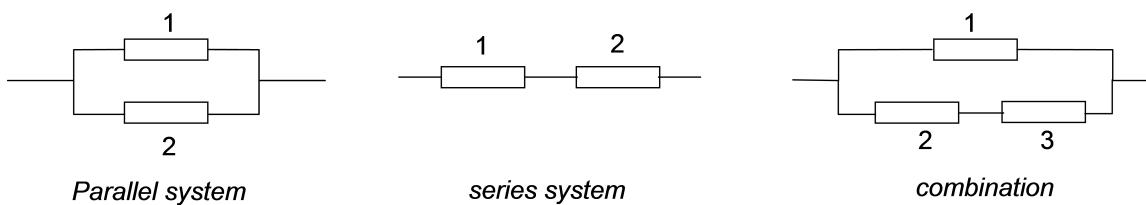


Figure 5.13: Schematic view of a parallel system, series system and a combination of both. Components 1, 2 and 3 can be viewed as bridges. The systems fails if a passenger can not walk from left to right over the bridges.

In the Probabilistic Library, failure probabilities of components are combined using the equivalent planes method, which is described in the following section.

A theoretical description of the method, with a focus on the Hohenbichler method, is presented in [section A.3.1](#).

5.5.1 Equivalent planes

Design points can currently only be combined afterwards. In such case, the initial design points are calculated individually for each failure definition and then combined. This combination uses an equivalent plane M for each design point, which provides an approximation. The equivalent plane is defined as follows:

$$M_i(u) = \beta_i + \sum_j \alpha_{i,j} \cdot u_j \quad (5.48)$$

and

$$M_{\text{combined}}(u) = \begin{cases} \text{series} & \min_i M_i(u) \\ \text{parallel} & \max_i M_i(u) \end{cases} \quad (5.49)$$

where:

- M_i is the result of the equivalent plane of design point i
- M_{combined} is the result of the combination of a number of design points i
- β_i is the reliability index of design point i
- $\alpha_{i,j}$ is the alpha factor of variable j in design point i

5.5.2 Combine methods

5.5.2.1 Directional sampling

Directional sampling (see section 5.3.7) is applied on the combined model M_{combined} (see Equation (5.49)). The following settings are applied:

Design point method	Center of gravity
Directions	1000 - 10000
Convergence factor	0.1

5.5.2.2 Importance sampling

Depending on the combination method (series or parallel) the following algorithms are applied:

Importance sampling - Series

Importance sampling (see section 5.3.4) is applied on the combined model M_{combined} (see Equation (5.49)) in the following iterative way: the combined probability of the first $N + 1$ design points is the probability of the first N design points added with the contribution of design point $N + 1$, where it does not overlap with one of the previous design points (see Equation (5.50)).

$$P(X_{1\dots N+1}) = P(X_{1\dots N}) + P(X_{N+1}) \cap P(\overline{X_{1\dots N}}) \quad (5.50)$$

where:

-
- $P(X_{1\dots N})$ is the combined probability of the first N design points
 $P(X_{1\dots N+1})$ is the combined probability of the first $N + 1$ design points
 $P(\overline{X_{1\dots N}})$ is the complement of the combined probability of the first N design points
 $P(X_{N+1})$ is the probability of design point $N + 1$

Equation (5.50) can be rewritten as:

$$P(X_{1\dots N+1}) = P(X_{1\dots N}) + P(X_{N+1}) \cdot (1 - P(X_{1\dots N}|X_{N+1})) \quad (5.51)$$

The last term ($P(X_{1\dots N}|X_{N+1})$) is calculated using Importance Sampling (see section 5.3.4). The starting point of the Importance Sampling is set to the design point of X_{N+1} . The calculation is stopped when the remainder of the design points represent a small probability of failure.

Importance sampling - Parallel

The parallel combination uses Crude Monte Carlo (see section 5.3.3). The sampling space is limited to the area which can lead to failure, which could be very small. Only values for u_i are allowed for which the following condition is true:

$$\forall k, u_{\min} < u_j < u_{\min} < u_i \exists u_i | M_k(u) = 0 \wedge u_{\min} < u_i < u_{\max} \quad (5.52)$$

where:

- i is the index of the i^{th} variable
- j is any other variable index ($j \neq i$)
- k is the index of the design point
- u_{\min} is the computationally minimum possible value of u (-8)
- u_{\max} is the computationally maximum possible value of u (8)
- $M_k(u)$ is the equivalent plane of the i^{th} design point (see Equation (5.48))

5.5.2.3 Hohenbichler

The core of the Hohenbichler algorithm is the combination of two models M_{combined} (see Equation (5.49)). This combination is calculated using FORM (see section 5.3.10).

The result of this combination is a design point, which is represented as an equivalent plane when it needs to be combined with a third design point. This process is repeated until all design points have been combined.

First, the two most contributing design points are combined, followed by the remaining design point with the highest probability of failure, and so on. This iterative approach ensures that the error associated with representing intermediate design points is minimized.

5.6 Upscaling for systems with identical components

Upscaling refers to combining failure probabilities over "identical components". Upscaling is distinguished from the more generic system combine methods because the components being identical allows for a more convenient and efficient combination procedure. Identical in this case refers to the fact that the components have the same failure probability (i.e. the same reliability index β) and they are mutually correlated with the same correlation coefficient ρ :

$$\begin{aligned}\beta(Z_i) &= \beta \quad ; i = 1 \dots n_e \\ \rho(Z_i, Z_j) &= \rho \quad ; i \neq j\end{aligned}\tag{5.53}$$

where n_e is the number of components, ρ is the correlation coefficient ($\rho \geq 0$) and Z_i is the Z -function of component i . Note that in general, the components also have in common the underlying set of random variables and the associated α -values, but that is not a necessary condition for applying the method as described in this section.

Examples of when upscaling may be applied are the combining of failure probabilities at one time scale to a larger time scale and upscaling failure probabilities from a cross section of a defense segment to the longitudinal extent of the segment. Such applications are described in [section 5.6.3](#). The value of ρ first needs to be determined from system knowledge. For now, ρ is assumed to be known.

5.6.1 Computational procedure

The failure probability of this system can be computed by solving the following integral numerically:

$$P(F) = \int_v [1 - \{1 - \Phi(-\beta^*)\}^n] \phi(v) dv\tag{5.54}$$

In which ϕ is the standard normal density function and β^* is equal to:

$$\beta^* = \frac{\beta - v\sqrt{\rho}}{\sqrt{1 - \rho}}\tag{5.55}$$

5.6.1.1 Detailed explanation

The upscaling method makes use of linearized approximations of the Z -functions, as described in [section A.1](#). The estimated probability of failure of the system is therefore an approximation of the true probability of failure. Errors made in the approximation will depend on the system under consideration. In [section A.1](#) it was shown that linearised Z -functions can be described as follows:

$$Z_i = \beta - \alpha_1 U_{i1} - \dots - \alpha_n U_{in} \quad ; i = 1 \dots n_e\tag{5.56}$$

Furthermore, it was shown that the sum of the product of α -values and standard normally distributed U -values can be replaced by a single standard normally distributed U -variable:

$$Z_i = \beta - U_i \quad ; i = 1 \dots n_e\tag{5.57}$$

where U_i is a standard normally distributed variable and β is the reliability index of each individual Z -function. The value of β is considered to be known, i.e. it is determined by

the probabilistic computation techniques as described in section 5.3. This means β is a constant in equation (5.57) and the mutual correlation of the Z -functions is therefore entirely determined by the mutual correlation of the U -variables:

$$\rho(Z_i, Z_j) = \rho(U_i, U_j) = \rho \quad ; i \neq j \quad (5.58)$$

To describe a system that satisfies the relation of equation (5.58), variable U_i is written as a function of two independent standard normal random variables U_i^* and V :

$$U_i = U_i^* \sqrt{1 - \rho} - V \sqrt{\rho} \quad (5.59)$$

The variables $U_i^*, i = 1 \dots n$ are taken to be mutually independent:

$$\rho(U_i^*, U_j^*) = 0 \quad ; i \neq j \quad (5.60)$$



Note: in the equations above, $\rho > 0$. For $\rho = 0$, the Hohenbichler method in combination with the outcrossing approach should be used. The outcrossing method is discussed further on in this chapter.

To verify the applicability of equation (5.59) it needs to be shown that [1] U_i is standard normally distributed and [2] that the relation of equation (5.58) holds. To prove [1], we apply the following general rule (see, e.g. Grimmett and Stirzaker (1983)): If X and Y are independent normally distributed random variables, then $aX + bY$ is also normally distributed with a mean, μ , and standard deviation, σ , equal to:

$$\begin{aligned} \mu &= a\mu_X + b\mu_Y \\ \sigma &= \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2} \end{aligned} \quad (5.61)$$

Application of this rule on equation (5.59), where U_i^* and V are both normally distributed with mean 0 and standard deviation 1, gives:

$$\begin{aligned} \mu &= \sqrt{1 - \rho} \cdot 0 - \sqrt{\rho} \cdot 0 = 0 \\ \sigma &= \sqrt{(1 - \rho) \cdot 1 + \rho \cdot 1} = 1 \end{aligned} \quad (5.62)$$

Which proves that U_i is standard normally distributed. To prove [2], i.e. that the relation of equation (5.58) holds, equations (5.57) and (5.59) are combined:

$$Z_i = \beta - U_i^* \sqrt{1 - \rho} - V \sqrt{\rho} \quad ; i = 1 \dots n_e \quad (5.63)$$

The variable V in equation (5.63) is part of each Z -function. This creates the desired mutual correlation between the functions Z_i , $i = 1 \dots n_e$. To prove this, it needs to be shown that variables U_i and U_j , $i \neq j$, have a mutual correlation equal to ρ . The correlation between U_i and U_j is derived as follows:

$$\rho(U_i, U_j) = \frac{\text{cov}(U_i, U_j)}{[\sigma(U_i) \sigma(U_j)]} = \frac{\text{cov}(U_i, U_j)}{[1 \cdot 1]} = \text{cov}(U_i, U_j) \quad (5.64)$$

The covariance of U_i and U_j is equal to:

$$\begin{aligned}
 &= E[U_i U_j - \mu(U_i) \mu(U_j)] = E[U_i U_j] \\
 cov(U_i, U_j) &= E[(U_i^* \sqrt{1-\rho} - V\sqrt{\rho})(U_j^* \sqrt{1-\rho} - V\sqrt{\rho})] \\
 &= E[U_i^* U_j^* (1-\rho) - U_i^* V \sqrt{\rho(1-\rho)} - U_j^* V \sqrt{\rho(1-\rho)} + \rho V^2] \\
 &= E[0 - 0 - 0 + \rho V^2] = \rho E[V^2] = \rho Var(V) = \rho
 \end{aligned} \tag{5.65}$$

This proves that equation (5.63) describes a system of n_e components that are mutually correlated with a correlation coefficient ρ . The theorem of total probability is used to derive the probability of failure of this system:

$$P(F) = P(Z_1 < 0 \cup \dots \cup Z_n < 0) = \int_v P(Z_1 < 0 \cup \dots \cup Z_n < 0|v) \phi(v) dv \tag{5.66}$$

where $\phi(v)$ is the probability density function of the standard normal distribution. The probability that at least 1 component fails is equal to 1 minus the probability that none of the components fails. Equation (5.66) can therefore be rewritten as:

$$\begin{aligned}
 P(F) &= \int_v [1 - P(Z_1 \geq 0 \cap \dots \cap Z_n \geq 0|v)] \phi(v) dv \\
 &= \int_v [1 - P\{(Z_1 \geq 0|v) \cap \dots \cap (Z_n \geq 0|v)\}] \phi(v) dv
 \end{aligned} \tag{5.67}$$

For a given value of V , the individual failure probabilities of the Z -functions are mutually independent:

$$P[(Z_i < 0|v) \cap (Z_j < 0|v)] = P(Z_i < 0|v) P(Z_j < 0|v) \quad ; i \neq j \tag{5.68}$$

This can be easily verified from equation (5.59). If the value of V is given, equation (5.59) only contains one random variable: U_i^* . Since the U_i^* -values are mutually independent (see equation (5.60)), the Z -functions of equation (5.59) are mutually independent as well, which leads to the equality in equation (5.68). Substitution of equation (5.68) in equation (5.67) gives:

$$P(F) = \int_v \left[1 - \prod_{i=1}^{n_e} P(Z_i \geq 0|v) \right] \phi(v) dv \tag{5.69}$$

Because all components are identical, the following is true:

$$P(Z_1 \geq 0|v) = P(Z_2 \geq 0|v) = \dots = P(Z_n \geq 0|v) \stackrel{\text{def}}{=} P(Z \geq 0|v) \tag{5.70}$$

This changes equation (5.69) into:

$$P(F) = \int_v [1 - P(Z \geq 0|v)^{n_e}] \phi(v) dv \tag{5.71}$$

The conditional probability of the Z -function in the integral is equal to:

$$P(Z \geq 0|v) = P\left(\beta - U^* \sqrt{1-\rho} - v\sqrt{\rho} \geq 0\right) = P\left(U^* \leq \frac{\beta - v\sqrt{\rho}}{\sqrt{1-\rho}}\right) \quad (5.72)$$

Since v is a given constant, U^* is the only random variable in equation (5.72), and since U^* is standard normally distributed, the conditional probability is defined as:

$$P(Z \geq 0|v) = \Phi\left(\frac{\beta - v\sqrt{\rho}}{\sqrt{1-\rho}}\right) \stackrel{\text{def}}{=} \Phi(\beta^*) \quad (5.73)$$

where Φ is the standard normal distribution function and β is equal to:

$$\beta^* = \frac{\beta - v\sqrt{\rho}}{\sqrt{1-\rho}} \quad (5.74)$$

Equation (5.71) then changes into:

$$P(F) = \int_v [1 - \{1 - \Phi(-\beta^*)\}^n] \phi(v) dv \quad (5.75)$$

The right hand side of equation (5.75) can be computed by numerical integration over the standard normally distributed variable V . Since V is the only variable, the grid size of V can be chosen small without requiring significant computation time. The error from the numerical integration of equation (5.75) can therefore be made as small as desired. This means the only potentially significant error that is introduced in this method is related to the linearisation of the Z -function, which was necessary to derive equation (5.75).

5.6.1.2 Equivalent α -values

As with the Hohenbichler method, equivalent α -values can be computed for the component that represents the combination of n_e identical components. This is necessary in case the resulting component is used in subsequent combining procedures where α -values are required. A similar approach with perturbed u -values is used as in the Hohenbichler method. However, because in this special case the components are identical, this allows for some convenient simplifications that require less computation time.

Computational procedure

The method for deriving equivalent α -values for the systems with identical components is as follows:

[1] Apply the upscaling method of section 5.6.1 to n_e components with reliability index β and mutual correlation ρ to derive the reliability index β^e of the combined (upscaled) component.

[2] Apply the upscaling method of section 5.6.1 on n_e components with reliability index $\beta - \varepsilon\sqrt{\rho}$ and mutual correlation ρ to derive the reliability index $\beta^e(\varepsilon)$ of the combined (upscaled) component.

[3] Determine α_v :

$$\alpha_v = \frac{\beta^e(\varepsilon) - \beta^e}{\varepsilon} \quad (5.76)$$

[4] For all random variables $k = 1 \dots n$, determine the equivalent α -value, α_k^e :

$$\alpha_k^e = \sqrt{1 - \alpha_v^2} \frac{\alpha_k \sqrt{1 - \rho_k}}{\sqrt{1 - \rho}} + \alpha_v \frac{\alpha_k \sqrt{\rho_k}}{\sqrt{\rho}} \quad (5.77)$$

in which ρ_k is the correlation between the k^{th} random variable of an element with the corresponding (k^{th}) random variable in another element

[5] Normalise the equivalent α -values:

$$\alpha_{k;final}^e = \frac{\alpha_k^e}{\sqrt{\sum_{j=1}^n (\alpha_j^e)^2}} ; k = 1 \dots n \quad (5.78)$$

The equivalent α -values of the n variables for the combined n_e components are obtained with only two applications of the upscaling method (steps 1 and 2 above). This is very efficient, when compared to the Hohenbichler method, which needs to be repeated $2n + 1$ times to derive the equivalent α -values for combining only two components. Another difference with the Hohenbichler method is that the method for identical components can be applied for non-integer values of n_e ; this is not the case for the Hohenbichler method. The application with non-integer values of n_e is useful for example for upscaling the failure probability of a cross section to the failure probability of a dike section, as will be explained later on in this document.

Detailed explanation

Consider the system of n_e identical components as described in equation (5.56):

$$Z_i = \beta - \alpha_1 U_{i1} - \dots - \alpha_n U_{in} ; i = 1 \dots n_e \quad (5.79)$$

These components are combined according to the method as described in [section 5.6.1](#), resulting in a failure probability and associated reliability index β^e . The combined component can be described by:

$$Z^e = \beta^e - \alpha_1^e U_1 - \dots - \alpha_n^e U_n \quad (5.80)$$

The superscript "e" in this equation refers to the fact that these are equivalent values and functions. In order to derive the α -values of function Z^e , recall from [section A.1](#) that the α -values of a Z -function are related to the reliability index β^e as follows:

$$\frac{\partial \beta^e}{\partial \bar{u}_k} = \alpha_k^e ; k = 1 \dots n \quad (5.81)$$

In which \bar{u}_i is the mean of variable U_i . Note the value α_k^e represents the combined effect of variables $U_{1k} \dots U_{n_e k}$, and that these variables are mutually correlated. In order to determine α_k^e , these variables need to be split in an independent and mutually dependent part, similar

to the description in section 5.6.1. Consider for this purpose equation (5.79). The different U -variables within a single component are mutually uncorrelated, whereas corresponding U -values in different components are correlated. In formula:

$$\begin{aligned}\rho(U_{ij}, U_{ik}) &= 0 \quad ; j \neq k \\ \rho(U_{ik}, U_{\ell k}) &= \rho_k \quad ; i \neq \ell\end{aligned}\tag{5.82}$$

Each U -variable can therefore be split in a correlated and uncorrelated part:

$$U_{ik} = U_{ik}^* \sqrt{1 - \rho_k} + V_k \sqrt{\rho_k}\tag{5.83}$$

Furthermore, the variables U_{ik}^* , $i = 1 \dots n_e$, $k = 1 \dots n$ and V_k , $k = 1 \dots n$ are all taken to be mutually independent:

$$\begin{aligned}\rho(U_{ij}^*, U_{k\ell}^*) &= 0 \quad ; i \neq j \cup k \neq \ell \\ \rho(U_{ij}^*, V_k) &= 0 \\ \rho(V_j, V_k) &= 0 \quad ; j \neq k\end{aligned}\tag{5.84}$$

With this formulation, variables U_{ik} , $i = 1 \dots n_e$, $k = 1 \dots n$, automatically fulfill requirement (5.82), as can be shown in the same manner as was done below equation (5.60) in the previous section. Substituting equation (5.83) in equation (5.79) gives:

$$\begin{aligned}Z_i &= \beta - \alpha_1 (U_{i1}^* \sqrt{1 - \rho_1} + V_1 \sqrt{\rho_1}) - \dots - \alpha_n (U_{in}^* \sqrt{1 - \rho_n} + V_n \sqrt{\rho_n}) \quad ; i = 1 \dots n_e \\ &= \beta - \sum_{k=1}^n \alpha_k U_{ik}^* \sqrt{1 - \rho_k} - \sum_{k=1}^n \alpha_k V_k \sqrt{\rho_k} \quad ; i = 1 \dots n_e\end{aligned}\tag{5.85}$$

This equation can be replaced by:

$$Z_i = \beta - U_i^* \sqrt{1 - \rho} - V \sqrt{\rho} \quad ; i = 1 \dots n_e\tag{5.86}$$

In which:

$$\begin{aligned}\rho &= \sum_{k=1}^n (\alpha_k)^2 \rho_k \\ U_i^* &= \frac{1}{\sqrt{1-\rho}} \sum_{k=1}^n \alpha_k U_{ik}^* \sqrt{1 - \rho_k} \quad ; i = 1 \dots n_e \\ V &= \frac{1}{\sqrt{\rho}} \sum_{k=1}^n \alpha_k V_k \sqrt{\rho_k}\end{aligned}\tag{5.87}$$

The validity of this replacement can be easily verified by substituting the formulations of U_i and V of equation (5.87) into equation (5.86). Equation (5.86) is equivalent to equation (5.63) if and only if U_i^* and V are mutually independent standard normally distributed variables. The mutual independency can easily be shown since all components U_{ik}^* , $i = 1 \dots n_e$, $k = 1 \dots n$, and V_k , $k = 1 \dots n$ are mutually independent (see (5.84)).

To verify if U_i^* and V are standard normally distributed we apply the following general rule (see, e.g. Grimmett and Stirzaker (1983)): If X and Y are independent normally distributed random variables, then $aX + bY$ is also normally distributed with a mean, μ , and standard deviation, σ , equal to:

$$\begin{aligned}\mu &= a\mu_X + b\mu_Y \\ \sigma &= \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}\end{aligned}\tag{5.88}$$

Application of this rule on equation (5.87), where all components U_{ik}^* and V_k are normally distributed with mean 0 and standard deviation 1, gives:

$$\begin{aligned}\mu(U_i) &= \frac{1}{\sqrt{\rho}} \sum_{k=1}^n \alpha_k \cdot 0 \cdot \sqrt{1 - \rho_k} = 0 \\ \sigma(U_i) &= \sqrt{\left(\frac{1}{\sqrt{1-\rho}}\right)^2 \sum_{k=1}^n (\alpha_k)^2 \cdot (1)^2 \cdot (1 - \rho_k)} = \sqrt{\frac{1}{1-\rho} [\sum_{k=1}^n \alpha_k^2 - \sum_{k=1}^n \alpha_k^2 \rho_k]} \\ &= \sqrt{\frac{1}{1-\rho} [1 - \rho]} = \sqrt{1} = 1\end{aligned}\tag{5.89}$$

This shows that U_i^* and V in equation (5.86) are mutually independent standard normally distributed variables. Taking into account the formulation of the Z -function in equation (5.86), The equivalent coefficient α_k^e can now be derived as follows

$$\alpha_k^e = \frac{\partial \beta^e}{\partial \bar{u}_k} = \frac{\partial \beta^e}{\partial \bar{u}} \frac{\partial \bar{u}}{\partial \bar{u}_k} + \frac{\partial \beta^e}{\partial \bar{v}} \frac{\partial \bar{v}}{\partial \bar{u}_k}\tag{5.90}$$

where \bar{u} , \bar{v} and \bar{u}_k are the mean values of variables U^* , V and U_k . So, the derivation of coefficients α_k $i = 1 \dots n_e$ it comes down now to determining the four partial derivatives of equation (5.90). The first two can be determined directly from equation (5.87):

$$\begin{aligned}\frac{\partial \bar{u}}{\partial \bar{u}_k} &= \frac{\alpha_k \sqrt{1-\rho_k}}{\sqrt{1-\rho}} \\ \frac{\partial \bar{v}}{\partial \bar{u}_k} &= \frac{\alpha_k \sqrt{\rho_k}}{\sqrt{\rho}}\end{aligned}\tag{5.91}$$

The partial derivative of β^e to \bar{v} is determined numerically:

$$\alpha_v = \frac{\partial \beta^e}{\partial \bar{v}} \approx \frac{\beta^e(\varepsilon) - \beta^e}{\varepsilon}\tag{5.92}$$

In which $\beta^e(\varepsilon)$ is the reliability index of the upscaled system of n_e components, after perturbation of \bar{v} with a small value ε .

$$\beta^e(\varepsilon) = \Phi^{-1}(1 - P(Z^e(\varepsilon) < 0)) = \Phi^{-1}\left(1 - \bigcup_{i=1}^{n_e} P(Z_i(\varepsilon))\right)\tag{5.93}$$

In which function $Z_i(\varepsilon)$ is as follows:

$$\begin{aligned}Z_i(\varepsilon) &= \beta - U_i^* \sqrt{1-\rho} - (V + \varepsilon) \sqrt{\rho} \quad ; i = 1 \dots n_e \\ &= (\beta - \sqrt{\rho}\varepsilon) - U_i^* \sqrt{1-\rho} - V \sqrt{\rho} \quad ; i = 1 \dots n_e\end{aligned}\tag{5.94}$$

In other words: $Z_i(\varepsilon)$ is a Z -function with a reliability index β_{Z_ε} following:

$$\beta_{Z_\varepsilon} = \frac{\mu_Z}{\sigma_Z} = \frac{\beta - \sqrt{\rho}\varepsilon}{1} = \beta - \sqrt{\rho}\varepsilon\tag{5.95}$$

So $\beta^e(\varepsilon)$ is quantified by substituting equation (5.94) into equation (5.93) and subsequent application of the upscaling procedure of section 5.6.1. Subsequently, $\beta^e(\varepsilon)$ is substituted in

equation (5.92) in order to derive α_v , the partial derivative of β^e to \bar{v} . The next step is to derive the partial derivative of β^e to \bar{u} . This can be derived as follows:

$$\frac{\partial \beta^e}{\partial \bar{u}} = \sqrt{1 - \left(\frac{\partial \beta^e}{\partial \bar{v}} \right)^2} = \sqrt{1 - \alpha_v^2} \quad (5.96)$$

This can be explained as follows: the partial derivative of β^e to \bar{v} is the resulting α -value for the dependent part of the n_e components, represented by variable V . The partial derivative of β^e to \bar{u} is the resulting α -value for the independent part of the n_e components, represented by variable U^* . The sum of the squares of these α -values should be equal to 1.

Substitution of equations (5.91), (5.92) and (5.96) into equation (5.90) provides the requested equivalent α -values:

$$\alpha_k^e = \frac{\partial \beta^e}{\partial \bar{u}_k} = \sqrt{1 - \alpha_v^2} \frac{\alpha_k \sqrt{1 - \rho_k}}{\sqrt{1 - \rho}} + \alpha_v \frac{\alpha_k \sqrt{\rho_k}}{\sqrt{\rho}} \quad (5.97)$$

The Z -function of the resulting component from the upscaling procedure (equation (5.80)) needs to have a standard deviation equal to 1. This means the sum of the squares of the equivalent α -values should be equal to 1. Equation (5.97) guarantees that this is the case if all values of ρ_k are equal to either 0 or 1, which is generally the case for upscaling in time (i.e. slow varying random load variables and strength variables have an autocorrelation equal to 1, while fast varying random variables have an autocorrelation equal to 0). This can be deducted as follows:

$$\begin{aligned} \sum_{k=1}^n (\alpha_k^e)^2 &= \sum_{k=1}^n \left[(1 - \alpha_v^2) \frac{\alpha_k^2 (1 - \rho_k)}{1 - \rho} + 2\alpha_v \left(\sqrt{1 - \alpha_v^2} \right) \frac{\alpha_k \sqrt{1 - \rho_k}}{\sqrt{1 - \rho}} \frac{\alpha_k \sqrt{\rho_k}}{\sqrt{\rho}} + \alpha_v^2 \frac{\alpha_k^2 \rho_k}{\rho} \right] \\ &= \sum_{k=1}^n \left[(1 - \alpha_v^2) \frac{\alpha_k^2 (1 - \rho_k)}{1 - \rho} + \alpha_v^2 \frac{\alpha_k^2 \rho_k}{\rho} \right] \\ &= \frac{(1 - \alpha_v^2)}{1 - \rho} \sum_{k=1}^n \alpha_k^2 (1 - \rho_k) + \frac{\alpha_v^2}{\rho} \sum_{k=1}^n \alpha_k^2 \rho_k \\ &= \frac{(1 - \alpha_v^2)}{1 - \rho} (1 - \rho) + \frac{\alpha_v^2}{\rho} \rho = (1 - \alpha_v^2) + \alpha_v^2 = 1 \end{aligned} \quad (5.98)$$

Note: in the second step of this equation, the middle term is removed because it is equal to zero (since either $\rho_k = 0$ or $1 - \rho_k = 0$). In the fourth step, equation (5.87) is used. If not all values of ρ_k are equal to either 0 or 1, the sum of the squares of the equivalent α -values is not necessarily equal to 1. In that case, they have to be normalized as follows:

$$\alpha_k^e = \frac{\alpha_k^e}{\sqrt{\sum_{j=1}^n (\alpha_j^e)^2}}; k = 1 \dots n \quad (5.99)$$



5.6.2 Techniques for time and space dependent processes

In this section, aspects of upscaling in time and space are addressed.

The techniques described in the previous sections all deal with system analysis of a discrete number of components which may represent dike sections, wind directions, etc. In some applications, however, Z is a function of space and time, which means in principle the number of components is infinite. This is schematically depicted in Figure 5.14. In the left panel, Z is

a time-dependent function and failure potentially can occur at any time. On the right, Z is a function of space, and failure can occur at any location.

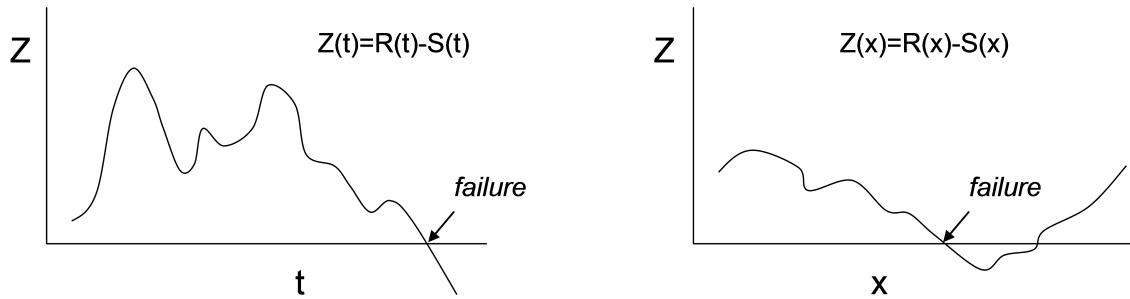


Figure 5.14: Stochastic variation of the Z -function in time (left) and space (right)

This section describes some approaches to deal with these type of continuous descriptions of Z -functions.

5.6.2.1 Poisson counting process

The Poisson counting process describes the probability of occurrence of n events, where a single event generally refers to an upcrossing or downcrossing of a threshold value. With respect to failure, the downcrossing of the threshold $Z = 0$ is most relevant. In a Poisson process it is assumed that for small values of Δt [a] the occurrence of an event in an interval $[t, t + \Delta t]$ is proportional to Δt and [b] the probability of occurrence of two events occurring in $[t, t + \Delta t]$ is negligible. This means for small values of Δt , the probability of an event occurring in $[t, t + \Delta t]$ is approximately equal to:

$$P(1 \text{ event during } [t, t + \Delta t]) \approx \nu \Delta t \quad (5.100)$$

In this equation, ν is the 'intensity' of the Poisson process. This is the single parameter that describes the Poisson process. Define $N(t)$ as the number of events occurring in the time interval $[0, T]$. For a Poisson process the probability distribution of $N(t)$ is:

$$P(N(t) = n) = \frac{(\nu t)^n e^{-\nu t}}{n!} \quad (5.101)$$

The time interval between two subsequent events is also a random variable and it is exponentially distributed. So if t_1 is the time interval between two events, then:

$$P(T_1 \leq t_1) = 1 - e^{-\nu t_1} \quad (5.102)$$

The assumption of a Poisson process is often used to translate exceedance frequencies into exceedance probabilities. Suppose ν is expressed as "number of events per year". In that case ν is the annual frequency of occurrence. Then, according to equation (5.102), the annual probability of occurrence is equal to:

$$P(T_1 \leq 1) = 1 - e^{-\nu} \quad (5.103)$$

This shows the relation between probability and frequency in case of a Poisson process. An event can be for instance the exceedance of a threshold level x , for load variable X . In that

case, equation (5.103) can be applied to translate the annual frequency of exceedance of threshold x into the annual probability of exceedance of threshold x , or vice versa.

In the description above, ν was assumed to be time-independent. If this is not the case, equation (5.101) changes into:

$$P(N(t) = n) = \frac{\left(\int_0^t \nu(\tau) d\tau\right)^n e^{-\int_0^t \nu(\tau) d\tau}}{n!} \quad (5.104)$$

5.6.2.2 Outcrossing

If an event refers to failure in a continuous process, i.e. the downcrossing of threshold $Z = 0$ in Figure 5.14, then the outcrossing rate is defined as:

$$\nu = \lim_{\Delta t \downarrow 0} \frac{P[Z(t) \geq 0 \cap Z(t + \Delta t) < 0]}{\Delta t} \quad (5.105)$$

Note that the numerator in this equation is the probability that failure occurs in time interval $[t, t + \Delta t]$. The rate ν is similar to the one defined in the previous section and can also be time-dependent: $\nu=\nu(t)$. Assume for the moment that ν is a constant, i.e. independent of time. The probability that failure occurs in an interval $(0, T]$, given the fact that no failure occurs at $t = 0$, is then equal to:

$$P \left[\min_{t \in (0, T]} \{Z(t)\} < 0 \mid Z(0) \geq 0 \right] = 1 - e^{-\nu T} \quad (5.106)$$

Note that this probability of failure is described by an exponential distribution function. The exponential distribution is by definition the distribution which describes failure probabilities for processes with a constant failure rate (see e.g. Grimmett and Stirzaker (1983)). Figure 5.15 shows an example of an exponential distribution function. In this figure the failure rate, ν , is taken equal to 1, which makes this a standard exponential distribution function.

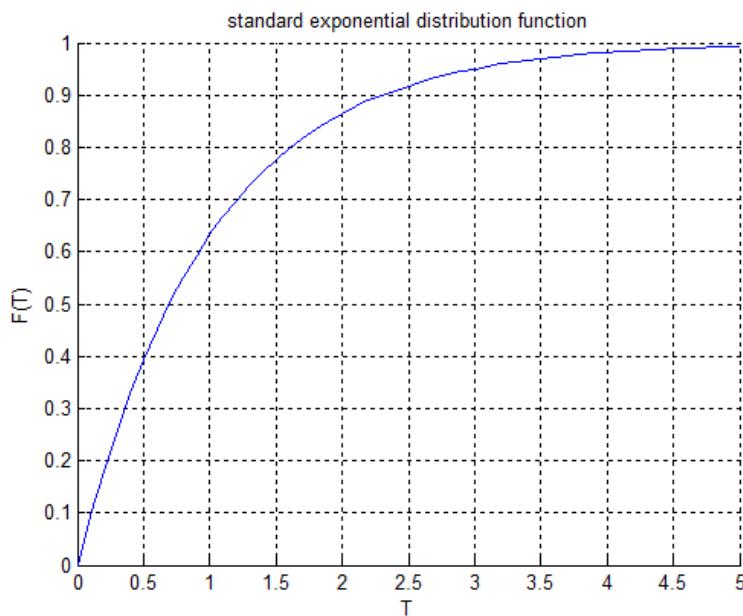


Figure 5.15: Standard exponential distribution function: $F(T) = 1 - \exp(-T)$.

The probability that no failure occurs in an interval $(0, T]$, i.e. $t = 0$ included, given the fact that no failure occurs at $t = 0$, is equal to:

$$P \left[\min_{t \in (0, T]} \{Z(t)\} \geq 0 \mid Z(0) \geq 0 \right] = e^{-\nu T} \quad (5.107)$$

The probability that no failure occurs in an interval $[0, T]$, i.e. $t = 0$ included, is then equal to:

$$P \left[\min_{t \in [0, T]} \{Z(t)\} \geq 0 \right] = [1 - P_F(0)] e^{-\nu T} \quad (5.108)$$

In which $P_F(0)$ is the initial probability of failure (see below for more information on this probability), i.e. the probability that $Z < 0$ at $t = 0$. The probability, P_f , that failure occurs in an interval $[0, T]$ is equal to:

$$P_F(T) = P \left[\min_{t \in [0, T]} \{Z(t)\} < 0 \right] = 1 - [1 - P_F(0)] e^{-\nu T} \quad (5.109)$$

If the outcrossing rate, ν , and the initial failure probability, $P_F(0)$, are small, the probability of failure can be approximated by:

$$P_F(T) \approx P_F(0) + \nu T \quad (5.110)$$

This is an upper bound of the failure probability. In essence, this approximation "double counts" the probability of events in which two or more failures occur in the interval $[0, T]$. If ν and $P_F(0)$ are small, the probability of two or more failures occurring in the interval $[0, T]$ is negligible and therefore equation (5.110) is a good approximation in that case.

In the equations above, failure rate ν was assumed to be constant. If this is not the case, equation (5.109) changes into the following, more generic, equation:

$$P_F(T) = 1 - [1 - P_F(0)] \exp \left(- \int_0^T v(t) dt \right) \quad (5.111)$$

The equations above can also be used if Z is a function of space. In that case, t and T need to be replaced by x and X , where x represents distance, e.g. the longitudinal distance along a dike section.

Note: in Probabilistic Library, the outcrossing method is applied in time as well as in space. First, the probabilities of failure of the smallest "components" are computed with the probabilistic techniques for single components as described in section 5.3. The smallest component is e.g. a cross section of a flood defense (space) during a tidal period (time) for a single failure mechanism. So, the initial result of the probabilistic procedure is the probability that failure occurs at a certain cross section within the time-span of a tidal period for a single mechanism. This result will be used as $P_F(0)$ in the equations above, i.e. the initial failure probability. Subsequently the outcrossing approach is applied for upscaling the probability of failure (for the mechanism under consideration) from a cross-section to a dike section and from a tidal period to a year.



The failure rate $\nu(t)$ or $\nu(x)$ needs to be derived from spatial and temporal autocorrelations of the strength and load variables. This is described in more detail in [Deltaires \(2023\)](#). In general, functions $\nu(t)$ and $\nu(x)$ are too complex to solve equation (5.111) analytically, which means approximating techniques are required. The Probabilistic Library uses different outcrossing approaches for space and time because of mutual differences in autocorrelation structures.

Note that the component for which $P_F(0)$ is computed in Probabilistic Library has a "width" equal to the assumed breach width. This means a (slight) reduction in the length of the remainder of the dike section and hence a (slight) reduction in the computed failure probability. The assumed breach width depends on the mechanism under consideration. The "width" in time is taken equal to a tidal period. This has to do with the fact that the input statistics of random load variables like sea water level, river discharge or wind speed represent probabilities of the maximum value in a tidal period (see also [Deltaires \(2023\)](#)). These values are therefore suitable to represent the whole tidal period.

5.6.3 Spatial upscaling - from cross section to flood defence segment

In this section, upscaling in space is discussed. This type of upscaling is applicable to flood defence systems.

5.6.3.1 Computing the failure probability

The spatial upscaling technique as described in the current section is done over homogeneous reaches of the flood defense. Homogenous in this case means the statistical characteristics remain constant. It is therefore relevant that the flood defence system is divided into segments for which the assumption of homogeneity is valid. So, if a dike segment is inhomogeneous, it needs to be split up into smaller, homogenous, segments.

Spatial upscaling is subject to a concept known as the length effect. The length effect essentially has to do with the increase in failure probability when going from a cross-section to a longitudinal segment and from a single segment to a flood defense system (interconnected segments). That is, the length effect refers to the effect that an increase in length has on the probability of failure. Note that this effect is also present when upscaling over time; the failure probability will increase as the considered time period increases.

The mathematical description of the length effect is the ratio of the failure probability of the larger length to that of the shorter. For the upscaling from cross-section to longitudinal segment (assuming statistical homogeneity!) this would be as follows:

$$\text{Length effect} = \frac{P_{f,segment}}{P_{f,cross-section}} \quad (5.112)$$

where $P_{f,segment}$ refers to the failure probability of the longitudinal segment and $P_{f,cross-section}$ refers to the failure probability of the cross section within that longitudinal segment. To derive the ratio of equation (5.112), a notion of the spatial correlation within the segment is required, for each random variable, X , involved. In Probabilistic Library this correlation is described with the following model:

$$\rho(\Delta y) = \rho_x + (1 - \rho_x) \exp \left[- \left(\frac{\Delta y^2}{d_x^2} \right) \right] \quad (5.113)$$

where ρ is the correlation between two locations within the segment ($\rho \geq 0$), Δy is the distance between these two locations, ρ_x is the residual correlation length of variable X and d_x is the spatial correlation length of variable X . Parameter d_x determines how quickly the correlation of variable X decreases over distance and ρ_x is the minimum correlation of

variable X between two locations of the same (homogeneous) segment. The parameters d_x and ρ_x need to be determined for each variable X , based on a combination of measurements and expert judgement.

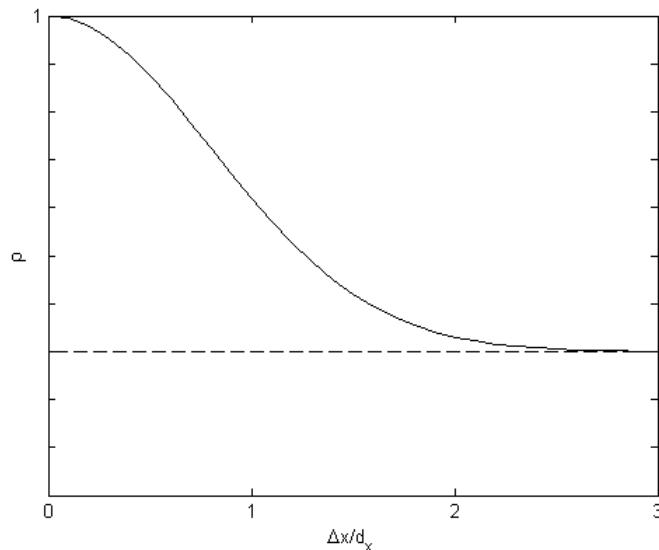


Figure 5.16: Autocorrelation function, correlation within a dike section; in this picture, the correlation ρ is visualized against Δx , made non-dimensional by the correlation distance d_x .

The correlation model of equation (5.113) and Figure 5.16 in principle is applied for each strength variable (load variables can generally be assumed to have correlation 1 within a single segment). This results in a similar model for the Z -function, i.e. in values d_Z and ρ_Z :

$$\rho(\Delta y) \approx \rho_Z + (1 - \rho_Z) \exp \left[- \left(\frac{\Delta y^2}{d_Z^2} \right) \right] \quad (5.114)$$

The parameters d_Z and ρ_Z can be derived as follows:

$$\rho_Z = \sum_{i=1}^n \alpha_i^2 \rho_i \quad (5.115)$$

$$\frac{1}{d_Z^2} = \frac{1}{1 - \rho_Z} \sum_{i=1}^n \alpha_i^2 (1 - \rho_i) \frac{1}{d_i^2} \quad (5.116)$$

In which:

- d_i is the correlation length of random variable i
- ρ_i is the residual correlation length of random variable i
- α_i is the influence coefficient of random variable i

Note that coefficients $\alpha_1, \dots, \alpha_n$ are determined in the probabilistic computation for a "representative" cross-section within the flood defence segment. For this purpose the probabilistic techniques for a single component are used (see section 5.3).

To derive the probability of failure of a dike segment, the segment is divided into components of equal length ΔL . The number of components is equal to:

$$n_e = \frac{L}{\Delta L} \quad (5.117)$$

where L is the length of the dike segment. The probability of failure for the entire dike segment is then equal to:

$$P_{f,segment} \approx (1 + n_e) P_{f,cross-section} = \left(1 + \frac{L}{\Delta L}\right) P_{f,cross-section} \quad (5.118)$$

This means the continuous process, in which failure can occur at any location along the dike is now replaced by a discrete process in which the dike segment is composed of a finite number of components, each of which has a failure probability that is equal to the probability of failure of a cross-section. This simplification/approximation is only valid for a well selected value of ΔL . If we assume that the spatial variation of Z is a Gaussian ergodic process (i.e. $\rho_Z = 0$), the length ΔL should be taken equal to:

$$\Delta L = d_Z \sqrt{\pi} / \beta \quad ; \text{if } \rho_Z = 0 \quad (5.119)$$

where β is the reliability index as derived in the probabilistic computation for a cross-section (see [section 5.3](#)). The value of ΔL is a result of the outcrossing approach (see [section 5.6.2.2](#)) in which the spatial variation of Z is assumed to be a Gaussian ergodic process. The derivation of ΔL , as described in equation (5.119), is described in [Jongejan \(2012\)](#).

With the assumption of a Gaussian ergodic process, the failure probability of a dike segment of length L is approximately equal to (combine equations (5.118) and (5.119)):

$$P_{f,segment} \approx \left(1 + \frac{L\beta}{d_Z \sqrt{\pi}}\right) \Phi(-\beta) \quad ; \text{if } \rho_Z = 0 \quad (5.120)$$

If $\rho_Z > 0$, the assumption of a Gaussian ergodic process does not hold and an alternative solution is required. In that case, $\rho_Z > 0$ represents the part of the correlation function that does not contribute to the length effect, because it is the correlation that persists over the entire dike segment. In that case the Z -function is split in an ergodic part (with ρ approaching zero over long distances) and a non ergodic part (with ρ constant):

$$Z = \beta - v\sqrt{\rho} - u\sqrt{1-\rho} \quad (5.121)$$

where v is the non-ergodic constant and u is the ergodic stochastic process with:

$$\rho(\Delta y) = \exp \left[- \left(\frac{\Delta y^2}{d_Z^2} \right) \right] \quad (5.122)$$

where ρ is the correlation between two locations within the segment and Δy is the distance between two locations. Using the theorem of total probability, the failure probability of the flood defense segment can be described as follows:

$$P[Z < 0] = \int P[Z < 0|v] f_V(v) dv \quad (5.123)$$

where $f_V(v)$ is the standard normal density function. The conditional failure probability, $P[Z < 0|v]$, in equation (5.123) can be written as (see Jongejan, 2012):

$$P[Z < 0|v] = 1 - (1 - P(Z_{cross} < 0)) e^{-N_f}$$

$$N_f = \frac{L}{2\pi} e^{-\frac{\beta^{*2}}{2}} \frac{\sqrt{2}}{d_z} \quad (5.124)$$

$$\beta^* = \frac{\beta_{cross} - v\sqrt{\rho_z}}{\sqrt{1 - \rho_z}}$$

where Z_{cross} and β_{cross} are the Z -function and reliability index of the cross section and Φ is the standard normal distribution function. The combination of equations (5.123) and (5.124) provide the probability of failure for a flood defense segment. More details on the derivation of equations (5.123) and (5.124) can be found in Jongejan (2012). Equation (5.124) can be evaluated with high accuracy using for example numerical integration.

Note: in formula 5.120 the width of the mechanism is not taken into account. In Probabilistic Library the width of the mechanism is taken equal to ΔL . In the formula for N_F the length L is replaced by $L - \Delta L$. The idea behind this correction is that for stretches smaller than ΔL it is not possible to have an increase in failure probability as a result of the length effect.



Note: formula (5.124) is only valid for value of $\rho_z > 0$. For values of $\rho_z = 0$ the Hohenbichler method is used in Probabilistic Library in combination with the outcrossing approach.



5.6.3.2 Computing equivalent α -values

As stated in the previous section, the flood defence segment can be thought of to consist of identical components of n identical components of length ΔL . Upscaling to a dike section in essence is therefore the same as upscaling over n identical components. The last step in such an upscaling process, is the derivation of new equivalent α -values for the individual random variables, see section 5.6.1.1. The first step in this method is to determine the α -value of the correlated part of the Z -function of equation (5.121), i.e. variable V . This is done in the standard way by perturbing the mean value of V with a small value ε and quantifying the effect on the computed β -value of the dike section of a small perturbation (ε) in the mean value of V . The α -value of V is thus equal to

$$\alpha_v = \frac{\partial \beta_{section}}{\partial \bar{v}} \quad (5.125)$$

Equation (5.97) states that the equivalent value, α_k^e , of variable k can then be derived as follows:

$$\alpha_k^e = \sqrt{1 - \alpha_v^2} \frac{\alpha_k \sqrt{1 - \rho_k}}{\sqrt{1 - \rho_Z}} + \alpha_v \frac{\alpha_k \sqrt{\rho_k}}{\sqrt{\rho_Z}} \quad (5.126)$$

In which α_k is the α -value of variable k before upscaling and ρ_k is the correlation of variable k between two components. Since components in this case have length ΔL , this correlation is equal to (see equation (5.113)):

$$\rho_k = \rho_{kr} + (1 - \rho_{kr}) \cdot \exp \left[- \left(\frac{\Delta L}{d_k} \right)^2 \right] \quad (5.127)$$

where ρ_{kr} is the residual correlation length of variable k and d_x is the spatial correlation length of variable k .

A Probabilistic computation techniques for system reliability

A.1 Linearisation of Z -functions

Z -functions in e.g. flood risk analysis generally describe a combination of hydrodynamics and geotechnical processes. Due to the complexity of the Z -function it is sometimes practical to use linear approximations. The linearisation can result in significant reductions of the computation time. This is for instance the case with the probabilistic computation method FORM, (see [section A.2.6](#)) which is generally much faster than other probabilistic computation techniques such as Monte Carlo (see [section A.2.3](#)).

Another advantage of the linearisation is that it enables (semi-)analytical approaches to complex system analysis, that otherwise would not have been possible. Such an approach is for instance used in the 'Hohenbichler method' (see [section A.3.1](#)) which is applied to compute the total probability of failure of a system of components.

The disadvantage of the linearisation is of course the fact that it is an approximation of the Z -function, which means an error is likely to be introduced in the estimate of the failure probability. As long as this error is small compared to other modeling errors and uncertainties this poses no real problem, but this needs to be verified as much as possible.

The linearisation of the Z -function is generally applied in the U -space, in which the U -variables are independent standard normally distributed random variables. In other words: the function $Z(u)$ is linearized, where U is the vector of standard normally distributed variables U_1, \dots, U_n . The linear approximation of the Z -function has the following form:

$$Z_L = B + A_1 U_1 + \dots + A_n U_n \quad (\text{A.1})$$

The linearisation is done by taking the tangent of the Z -function in a selected location $U = u_d$. This means the A -values are chosen as follows:

$$A_i = \frac{\partial Z}{\partial u_i}(u_d) \quad ; i = 1 \dots n \quad (\text{A.2})$$

This linearisation process is depicted in [Figure A.1](#) and [Figure A.2](#).

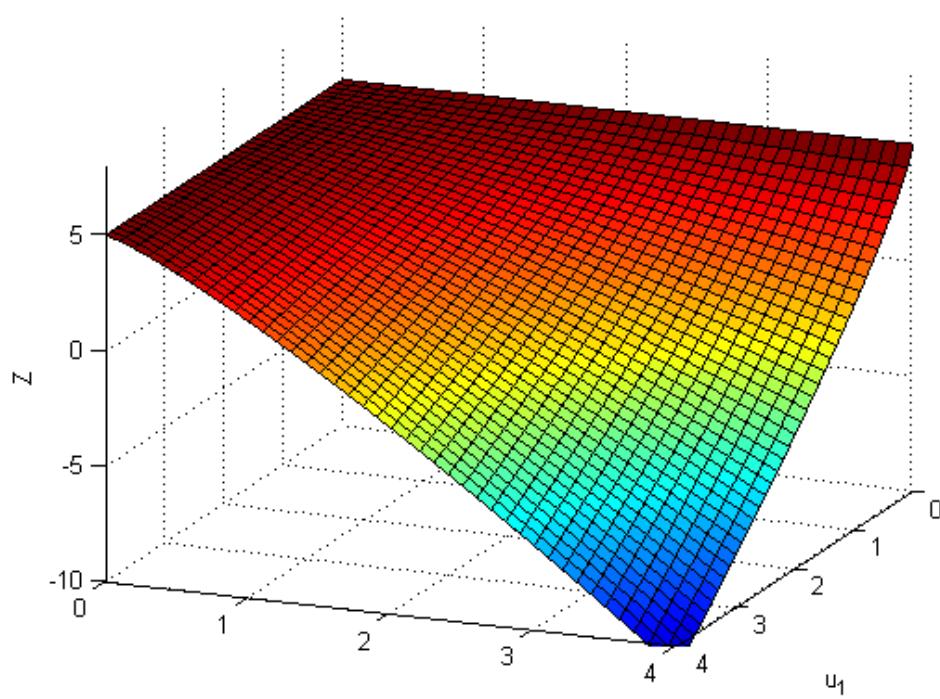


Figure A.1: Example of function $Z(U_1, U_2)$

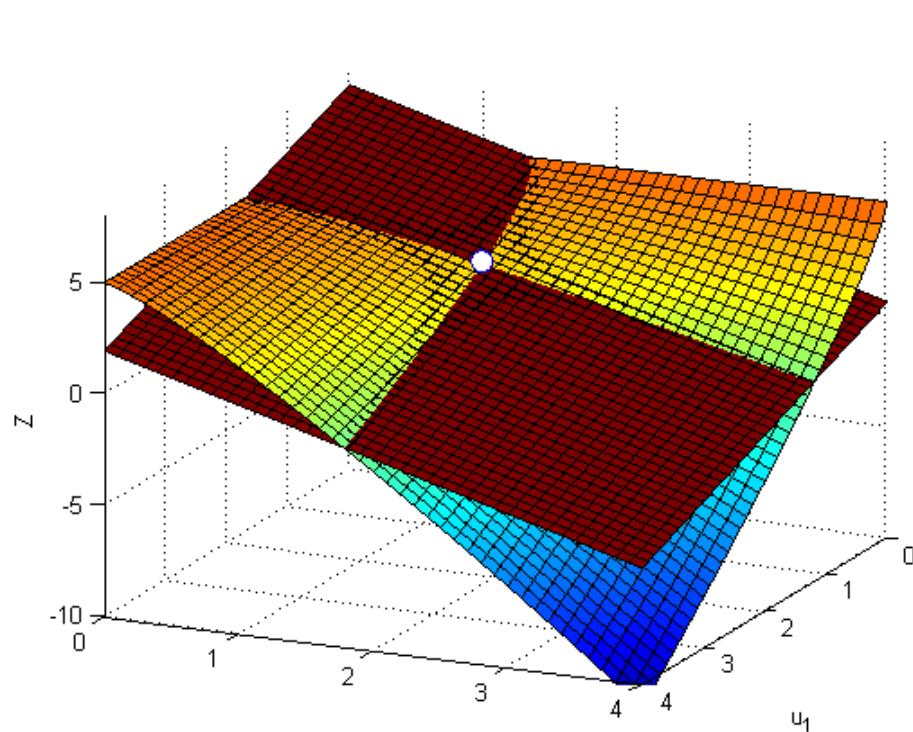


Figure A.2: Linearisation of the Z -function (dark red plane) of the Z -function of Figure A.1 in a selected location (white dot).

Clearly, the linearised Z -function is different from the actual Z -function. This means an error

will be introduced in the estimation of the probability of failure, $P(Z < 0)$. To reduce this error as much as possible, the linearisation is generally done in the design point, u_d . This is the location on the hyperplane $Z = 0$ with the highest probability density. The method FORM, as described in section A.2.6 is based on this principle.

Generally, the main objective is to compute the probability of failure, i.e. $P(Z < 0) \approx P(Z_L < 0)$. In that case, the right hand side of equation (A.1) can be multiplied or divided by a constant. If this constant is taken to be the norm of vector $A = (A_1, \dots, A_n)$ the linear Z -function has the following form:

$$Z_L = \beta + \alpha_1 U_1 + \dots + \alpha_n U_n = \beta + \sum_{i=1}^n \alpha_i U_i \quad (\text{A.3})$$

In which:

$$\beta = \frac{B}{\|A\|}; \quad \alpha_i = \frac{A_i}{\|A\|}, i = 1 \dots n; \quad \|A\| = \sqrt{\sum_{i=1}^n A_i^2} \quad (\text{A.4})$$

The norm of vector $\alpha = (\alpha_1, \dots, \alpha_n)$ is then equal to 1:

$$\sqrt{\sum_{i=1}^n \alpha_i^2} = 1 \quad (\text{A.5})$$

This means the linearised Z -function has been normalised. Since the U -variables are independent standard normally distributed values, this means:

$$\sum_{i=1}^n \alpha_i U_i \sim N(0, 1) \quad (\text{A.6})$$

In other words: the sum of the product of α -values and U -variables, $\sum \alpha_i U_i$, is standard normally distributed. This means that in order to compute $P(Z_L < 0)$, $\sum \alpha_i U_i$ can be replaced by a single standard normally distributed variable U^* :

$$Z_L = \beta + U^* \quad (\text{A.7})$$

Note that, since the density function of U^* is symmetric around $U^* = 0$, Z_L can also be described as follows:

$$Z_L = \beta - U^* \quad (\text{A.8})$$

In equation (A.7), failure occurs if $Z_L < 0$, i.e. if $U^* < -\beta$. The probability that this occurs is equal to $\Phi(-\beta)$, where Φ is the standard normal distribution function and β is the reliability index which was introduced in section 5.2. While β is an indicator for the probability of failure, the α -values are indicators for the relative importance of the associated random variables, as will be shown below. From equation (A.3) it can be seen that:

$$P(Z_L < 0) = P\left(\beta + \sum_{i=1}^n \alpha_i U_i < 0\right) \quad (\text{A.9})$$

If we increase the mean of variable U_i (\bar{u}_i) with a small value ε_i this will have an effect on the probability of failure. The magnitude of this effect is an indicator of the relative importance of variable U_i . For this purpose, define the random variable U'_i as follows:

$$U'_i = U_i + \varepsilon_i \quad (\text{A.10})$$

Since U_i is standard normally distributed, U'_i is normally distributed with mean ε_i and standard deviation 1. Subsequently, U_i in equation (A.3) is replaced by U'_i , resulting in a new Z -function Z'_L :

$$\begin{aligned} Z'_L &= \beta + \alpha_1 U_1 + \dots + \alpha_i U'_i + \dots + \alpha_n U_n \\ &= \beta + \alpha_1 U_1 + \dots + \alpha_i (U_i + \varepsilon_i) + \dots + \alpha_n U_n \\ &= (\beta + \alpha_i \varepsilon_i) + \alpha_1 U_1 + \dots + \alpha_i U_i + \dots + \alpha_n U_n \end{aligned} \quad (\text{A.11})$$

So the perturbation of the mean of variable U_i results in a new Z -function with reliability index β' instead of β , with:

$$\beta' = \beta + \alpha_i \varepsilon_i \quad (\text{A.12})$$

This means:

$$\frac{\partial \beta}{\partial \bar{u}_i} = \frac{\partial \beta}{\partial \varepsilon_i} = \frac{\beta' - \beta}{\varepsilon_i} = \frac{(\beta + \alpha_i \varepsilon_i) - \beta}{\varepsilon_i} = \frac{\alpha_i \varepsilon_i}{\varepsilon_i} = \alpha_i \quad (\text{A.13})$$

In other words: α_i is a measure of the sensitivity of reliability index β to changes in the mean value of variable U_i . This also means α_i is a measure of the sensitivity of the probability of failure to changes in the mean value of variable U_i . This information is used in the Hohenbichler method for combining probabilities of components in a system (see section A.3.1).

Note: as stated before, linearized Z -functions are the basis for various computation techniques that are explained in the following sections. A full understanding of this linearisation process and the meaning of α -values and β is therefore essential for further reading of this document.



A.2 Failure probability for a single component

In this section, aspects of computing the failure probability for a single component are addressed.

A.2.1 Introduction

This section describes in detail the computation techniques presented in Table A.1.

Table A.1: Computation techniques available in the Probabilistic Library for the computation of failure probability of a cross section of a longitudinal segment.

Method	Variant
Numerical integration	–
Monte Carlo	Crude Importance sampling Directional sampling
FORM (First Order Reliability Method)	–

A.2.2 Numerical Integration

Numerical integration solves equation (5.4) by discretizing the random variables $X_1 \dots X_n$. Each variable is discretized over a range that is relevant for failure, and subsequently each combination of discretized values of the X -variables is used to compute the limit state function. The probabilities of all the combinations that lead to $Z < 0$ are summed, which provides the estimate of the overall probability of failure. This summation can be written as follows:

$$\hat{P}_f = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \cdots \sum_{i_n=1}^{m_n} 1_{[Z<0]} f_X(x_{0,1} + (i_1 - 0.5)\Delta x_1, x_{0,2} + (i_2 - 0.5)\Delta x_2, \dots, x_{0,n} + (i_n - 0.5)\Delta x_n) \Delta x_1 \Delta x_2 \dots \Delta x_n \quad (\text{A.14})$$

where:

- \hat{P}_f = Estimated probability of failure
- $1_{[Z<0]}$ = Indicator function, equal to 1 for $Z < 0$, equal to 0 for $Z \geq 0$
- $x_{0,k}$ = Lower range limit for the k^{th} variable
- Δx_k = Interval width of the k^{th} variable
- m_k = Upper bound of k such that $x_{0,k} + m_k \Delta x_k$ is the upper bound of the k^{th} variable

In equation (A.14), for each variable X_k an equidistant grid with step size Δx_k is used, but non-equidistant grids can also be used in numerical integration. Figure A.3 presents a schematic view of the method, for an example of two random variables X_1 and X_2 . A 2-dimensional grid is defined and the Z -function is evaluated at the centre of the grid cells. Red grid points indicate failure ($Z < 0$), green grid points indicate no failure ($Z \geq 0$). The total probability of failure (see equation (A.14)) is estimated as follows: multiply the probability density of the grid cells in the failure domain (red dots) with the size of the grid cells ($\Delta x_1 \times \Delta x_2$) and take the sum of these probabilities.

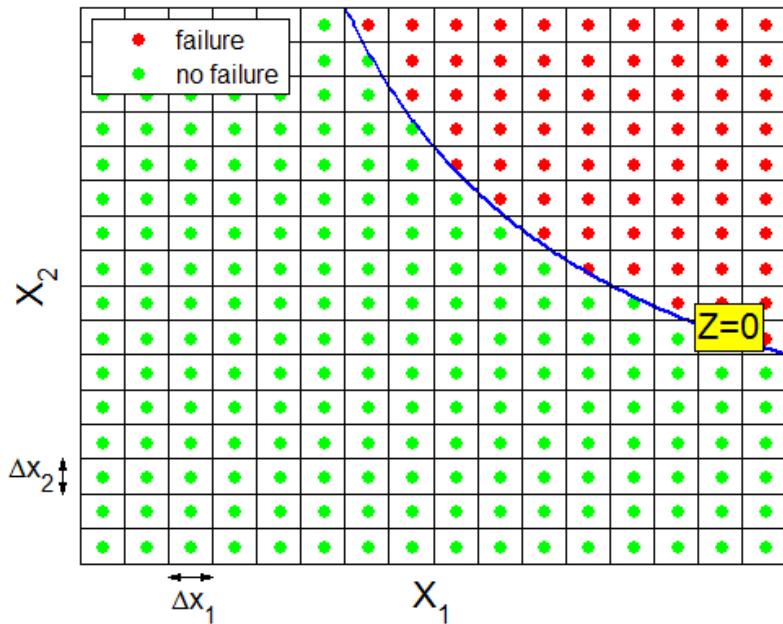


Figure A.3: Schematic view of the method of numerical integration for an example of two random variables. A 2-dimensional grid is defined and the Z -function is evaluated at the centre of the grid cells. Red grid points indicate failure ($Z < 0$), green grid point indicate no failure ($Z \geq 0$).

Like every probabilistic estimation technique, the result of the numerical integration procedure will be an approximation of the actual probability of failure. The errors that are introduced in this method are caused by the following assumptions and approximations:

Each grid cell is assumed to be entirely situated in the failure domain or entirely situated outside the domain of failure domain. In reality, grid cells can be partly in the failure domain as can be seen in Figure A.3.

The probability density is assumed to be constant over the entire grid cell. The domain of potential outcomes of the random variables may not be entirely covered. In the implementation of the procedure, it may be beneficial to transform the X -variables to standard normally distributed U -variables (see section 2.3). One of the benefits of working in the U -space is that the U -variables are independent, which simplifies equation (A.14) as follows:

$$\hat{P}_f = \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} 1_{[Z<0]} \phi(u_{0,1} + (i_1 - 0.5) \Delta u_1) \cdots \phi(u_{0,n} + (i_n - 0.5) \Delta u_n) \Delta u_1 \cdots \Delta u_n \quad (\text{A.15})$$

where:

- \hat{P}_f = Estimated probability of failure
- ϕ = standard normal density function
- $u_{0,k}$ = Lower range limit for the k^{th} variable, in the U -space
- Δu_k = Interval width of the k^{th} variable, in the U -space
- m_k = Upper bound of k such that $u_{0,k} + m_k \cdot \Delta u_k$ is the upper bound of the k^{th} variable

A.2.3 Crude Monte Carlo

Crude Monte Carlo sampling refers to the repeated sampling of the variables from the multivariate probability distribution function $f_X(x)$ (or, if the variables are mutually independent, sampling from the respective distribution functions $f_{X_1}(x_1), \dots, f_{X_n}(x_n)$). A single sample x_i refers to a vector of length n , where n is the number of random variables. For each sample x_i , the resulting value of limit state function $Z(x_i)$ is computed. The probability of failure is estimated as the ratio of samples for which $Z(x_i) < 0$, N_f , to the total number of samples, N :

$$\hat{P}_f = \frac{N_f}{N} = \frac{\sum_{i=1}^N I(Z(x_i))}{N} \quad (\text{A.16})$$

where I is the indicator function, which is equal to unity when $Z < 0$, equal to zero when $Z \geq 0$. Figure A.4 shows a schematic view of the procedure for an example with two random variables X_1 and X_2 . Each dot represents a sampled pair (x_1, x_2) . Red grid points indicate failure ($Z < 0$), green grid point indicate no failure ($Z \geq 0$). The estimated probability of failure is equal to the number of the red dots divided by the total number of dots.

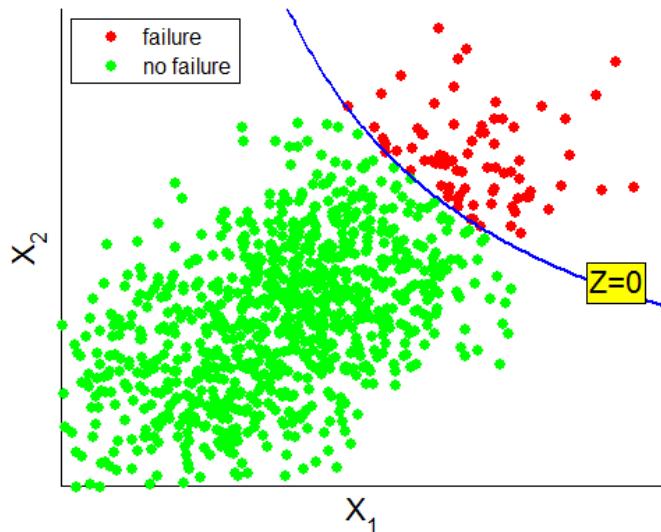


Figure A.4: Schematic view of Monte Carlo sampling for an example with two random variables. Each dot represents a sampled pair (x_1, x_2) . Red dots indicate failure ($Z < 0$), green dots indicate no failure ($Z \geq 0$).

The required number of samples, N , to provide a reliable estimate of the probability of failure depends on the actual failure probability P_f and on the acceptable error in the estimate of P_f . Additionally, it depends on the acceptable probability that the real error is within the accepted range. This is because even though taking a large number of samples will most likely result in small errors (law of large numbers), it can never be fully guaranteed due to the random character of the Monte Carlo sampling. However, it is possible to take N large enough to guarantee with for example 95% or 99% certainty that the error in the estimate is within the acceptable range. This probability, P_k , can be expressed as:

$$P_k = \Phi(k) - \Phi(-k) \quad ; k > 0. \quad (\text{A.17})$$

where Φ represents the standard normal distribution function, and k represents a sort of reliability index that the error is within the accepted range. The relation between k and P_k is schematically depicted in Figure A.5.

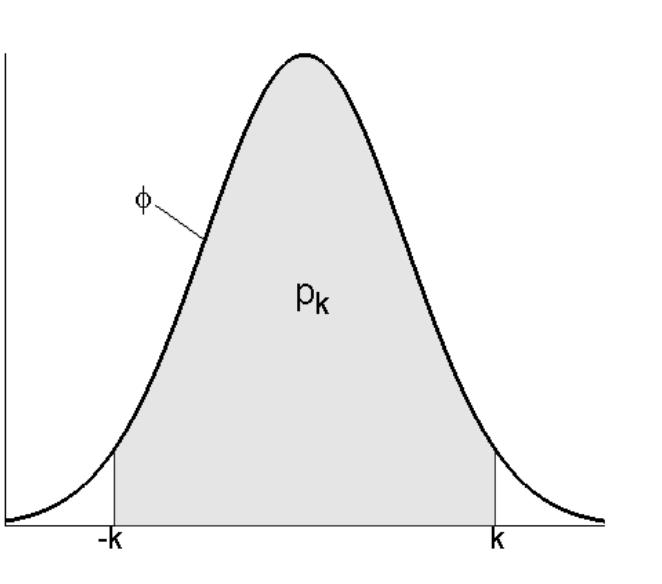


Figure A.5: Relation between k and P_k according to equation (A.18). Function ϕ is the standard normal density function

For example, a probability of 95% corresponds with a k -value of 1.96, because 95% of samples from a standard normal distribution function have a value between -1.96 and 1.96 . In formula, k is defined as:

$$k = \Phi^{-1} \left(\frac{1 + P_k}{2} \right) \quad (\text{A.18})$$

where P_k is the desired probability that the actual error is within the defined acceptable range. The required number of samples N can be estimated with the following formula (Melchers, 2002):

$$N = \frac{k^2}{\varepsilon^2} \left(\frac{1 - P_f}{P_f} \right) \quad (\text{A.19})$$

where ε is the acceptable relative error in the estimate of P_f :

$$\varepsilon = \left(\frac{|\hat{P}_f - P_f|}{P_f} \right) \quad (\text{A.20})$$

where \hat{P}_f is the Monte Carlo estimator of failure probability P_f .

Note that the required number of samples depends on the failure probability, which is not known in advance. Therefore, an estimate of the order of magnitude of the failure probability must be assumed, which can subsequently be revised during the Monte Carlo sampling procedure.

Table A.2 shows the required number of samples for combinations of ε and P_f . The value of k in this example is taken equal to 1.96. The numbers from this table show that, taking into account that each sample involves an evaluation of the Z -function, crude Monte Carlo is a rather inefficient method (i.e. a large number of Z -function evaluations is required) especially for estimating small failure probabilities.

Table A.2: Required number of samples with crude Monte Carlo for combinations of the acceptable relative error ε and actual probability of failure P_f . The value of k in equation (A.18) is taken equal to 1.96.

	ε		
P_f	0.10	0.05	0.01
10^{-2}	$4 \cdot 10^4$	$2 \cdot 10^5$	$2 \cdot 10^7$
10^{-3}	$4 \cdot 10^5$	$2 \cdot 10^6$	$2 \cdot 10^8$
10^{-4}	$4 \cdot 10^6$	$2 \cdot 10^7$	$2 \cdot 10^9$
10^{-5}	$4 \cdot 10^7$	$2 \cdot 10^8$	$2 \cdot 10^{10}$

A.2.4 Monte Carlo Importance Sampling

In this section, aspects of importance sampling as an addition computational component to Monte Carlo type methods are addressed.

A.2.4.1 General description

Importance sampling is a method to increase the efficiency of the crude Monte Carlo method; that is, to decrease the number of samples and Z -function evaluations required to produce a reliable estimate of the failure probability. This is done by replacing the initial probability density, f_X , of the input variables by a more efficient one, h_X , in which "efficient" refers to the proportion of the samples which will result in failure. An increasing percentage of samples in the failure domain results in a reduction in the variance of the estimator of the failure probability, hence a smaller number of samples is required for a reliable estimate.

There are a number of ways in which importance sampling can be applied; two of these are described in this section. The first increases the variance of the density function, resulting in a higher likelihood that failure events are sampled. The second essentially shifts the density function towards the failure domain so that, again, the likelihood of a failure sample increases. These two methods are illustrated in Figure A.6; the left-hand side illustrates the concept of a shifting of the density function, and the right-hand side illustrates the concept of increased variance.

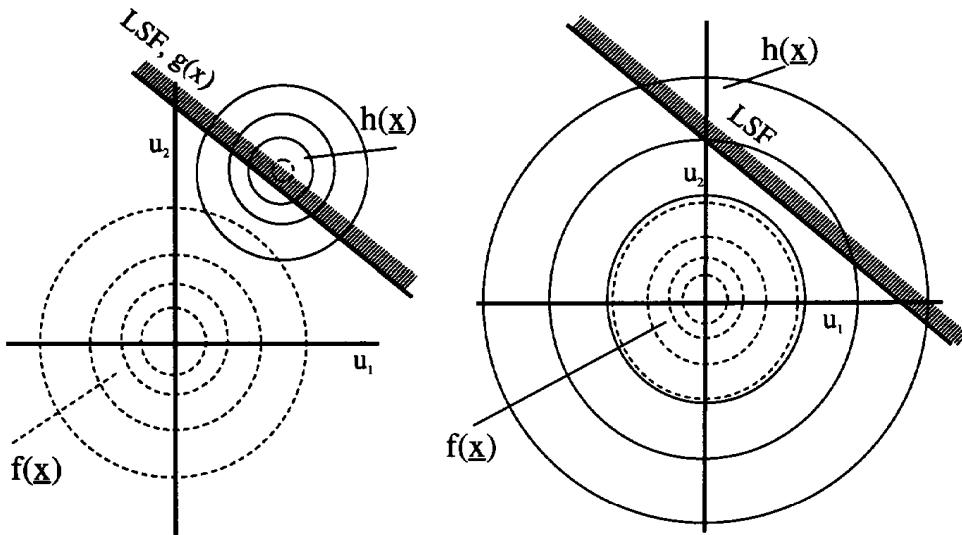


Figure A.6: Concept of importance sampling; left shows the concept of density shifting, right shows the concept of increased variance. The limit state function (LSF) is illustrated. Note that the LSF's are linear in this figure, but this is by no means a requirement for the applicability of importance sampling.

Because the sampling hasn't taken place from the initial distribution, the typical estimator of the failure probability (see equation (A.16)) needs to be corrected for this fact. This is done via the following formula:

$$\hat{P}_f = \frac{\sum_{i=1}^N I(Z(x_i)) \frac{f_X(x_i)}{h_X(x_i)}}{N} \quad (\text{A.21})$$

where \hat{P}_f is the estimated probability of failure, I is the indicator function (equal to unity when $Z < 0$, equal to zero when $Z \geq 0$), N is the total number of samples taken, f_X is the density function of x and h_X is the importance sampling density function.

Equation (A.21) can be explained by comparing it with equation (A.16), which describes the crude Monte Carlo method. In both equations, the indicator I is equal to one if the sampled vector x_i is in the failure domain and equal to 0 if x_i is outside the failure domain. In the crude Monte Carlo method, each sampled failure event scores a point and the more points scored, the higher the estimated probability of failure. In the importance sampling method it needs to be taken into account that the sampling of vector x_i was influenced by the fact that the density function was changed: $h_X(x)$ was applied instead of the real density function $f_X(x)$. This means the probability of sampling x_i was increased by a factor $c = h_X(x_i)/f_X(x_i)$. This manipulation in the density function needs to be compensated for in the "scoring". Therefore, a sampled event x_i in the failure domain does not score a full point, but "only" $1/c = f_X(x_i)/h_X(x_i)$.

So, the difference between equation (A.21) and (A.16) is the correction term f_X/h_X . This correction is necessary to make the estimate of P_f unbiased (provided h is well chosen) and accordingly that the error in the estimate can be made as small as desired by taking a sufficiently large number of samples, N . For importance sampling there is no simple generic error estimate like equation (A.19) for Crude Monte Carlo sampling, because the error estimate depends on the choice of $h_X(x)$. The efficiency of importance sampling therefore also depends strongly on the choice of $h_X(x)$. Prior knowledge of the problem under consideration

is therefore necessary to be able to define an efficient importance sampling method. Without such knowledge, there is even the potential danger that the important area for the limit state function (LSF) will be missed.

A.2.4.2 Implementation of the method of increased variance

The implementation of this general formula in the Probabilistic Library for the case of increased variance will now be described. Equations (A.22) through (A.27) show how the formula programmed in the Probabilistic Library can be derived from the general expression in equation (A.21). A random sampling of standard normal variables is first done; let us refer to these variables as U_1 , where U_1 is the vector containing all the variables. Subsequently, each u_1 -value is multiplied by a constant factor, a , and increased with a shift, b , to obtain a sample $u = b + a \cdot u_1$ (in the Probabilistic Library, a and b can differ per variable). Thus, the initial distribution of each u_1 -value is standard normal. The new set of variables, U , then has a normal distribution with a mean value equal to b and a standard deviation equal to a . The general form of the normal distribution is given below for reference.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (\text{A.22})$$

The ratio $f(u)/h(u)$ is then derived as follows, shown for the case of one variable:

$$f(u) = \frac{1}{\sqrt{2\pi(1)^2}} \exp\left[-\frac{(u-0)^2}{2(1)^2}\right] = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] \quad (\text{A.23})$$

$$h(u) = \frac{1}{\sqrt{2\pi a^2}} \exp\left[-\frac{(u-b)^2}{2a^2}\right] \quad (\text{A.24})$$

$$\frac{f(u)}{h(u)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right]}{\frac{1}{\sqrt{2\pi a^2}} \exp\left[-\frac{(u-b)^2}{2a^2}\right]} = a \frac{\exp\left[-\frac{u^2}{2}\right]}{\exp\left[-\frac{(u-b)^2}{2a^2}\right]} \quad (\text{A.25})$$

Writing u in terms of u_1 , the equation becomes:

$$\frac{f(u)}{h(u)} = a \frac{\exp\left[-\frac{u^2}{2}\right]}{\exp\left[-\frac{(b+au_1-b)^2}{2a^2}\right]} = a \frac{\exp\left[-\frac{u^2}{2}\right]}{\exp\left[-\frac{u_1^2}{2}\right]} \quad (\text{A.26})$$

Note that the one-variable example can easily be expanded to more variables, using the property that the u -values are independent, and hence their probability densities can be multiplied for the multi-variate case. The multi-variate form of equation (A.26) is:

$$\frac{f(u)}{h(u)} = \prod_k a_k \frac{\exp\left[-\frac{1}{2} \sum_k u_k^2\right]}{\exp\left[-\frac{1}{2} \sum_k u_{1k}^2\right]} \quad (\text{A.27})$$

where k runs over the total number of random variables. Equation (A.27) is the form in which the ratio is programmed in the Probabilistic Library.

A.2.5 Monte Carlo Directional Sampling

Directional Sampling (see, e.g. [Bjerager \(1988\)](#)), also referred to as Directional Simulation is a type of Monte Carlo method which aims to (strongly) reduce the number of samples in comparison with the Crude Monte Carlo method.

Directional sampling is depicted in Figure A.7. The first step in the method is to sample a direction in the standard-Normal space. This is done by sampling $\mathbf{u} = (u_1, u_2, \dots, u_n)$. For the i^{th} sample, $\mathbf{u}_i = (u_{1i}, u_{2i}, \dots, u_{ni})$, the directional unit vector θ_i is obtained by normalizing \mathbf{u}_i as described in Eq. A.28. The next step in the method is to determine, for the given direction θ_i , the value (or length) λ_i such that the limit state function evaluated at $\lambda_i \cdot \theta_i$ equals zero. The procedure to determine the length λ is described in the following subsection. The conditional failure probability for the i^{th} sample, P_i , is then calculated using the Chi-square (χ^2) distribution (see Eq. A.29). The χ^2_n distribution with n degrees of freedom is the distribution function of the sum of the squares of n independent standard normal random variables. Because we work in the standard normal space, the probability that $\|\mathbf{u}_i\| \leq \lambda_i$ is described by this distribution function. The final step is to estimate the failure probability as the mean of the conditional probabilities over all N sampled directions (Equation A.30).

$$\theta_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|} = (\bar{u}_1, \bar{u}_2, \dots, \bar{u}_n) \quad (\text{A.28})$$

$$P_i = 1 - \chi_n^2(\lambda_i^2) \quad (\text{A.29})$$

$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^N P_i \quad (\text{A.30})$$

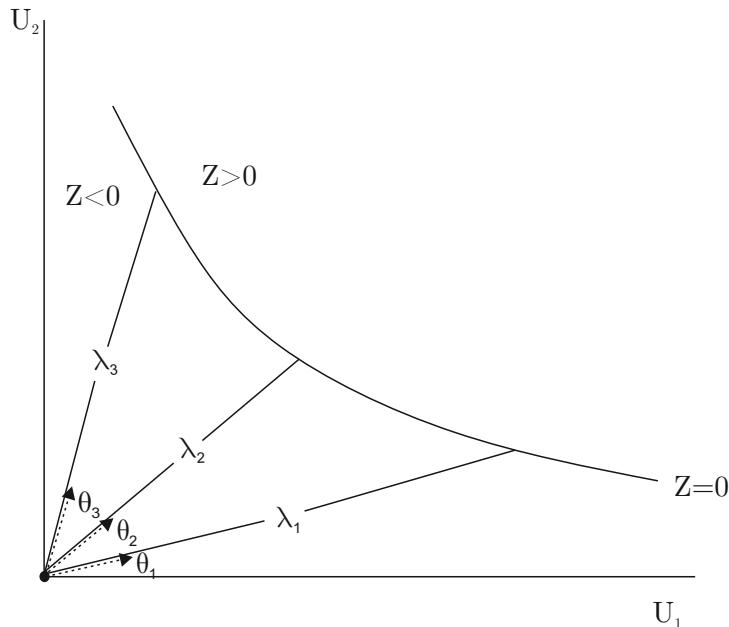


Figure A.7: Schematic view of the directional sampling method in the standard normal space. The distances λ_i to the limit state ($Z = 0$) are shown for unit-length directional vectors θ_i for three directional samples ($i = 1, 2, 3$).

Similar to other Monte Carlo methods, the outcome of the estimated probability of failure is a random variable and the error in the estimate can be made as small as possible by taking a sufficient number of samples. For directional sampling, the standard deviation, σ , of the estimated probability of failure can be quantified as follows (see, e.g. [Rosenblatt \(2011\)](#), [Melchers \(2002\)](#) page 84):

$$\sigma_{\hat{P}_f} = \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (P_i - \hat{P}_f)^2} \quad (\text{A.31})$$

From this equation, relative errors and confidence intervals can be estimated. Note that the error in the estimated failure probability is a random variable that approaches a normal distribution function, as N increases. This follows from the central limit theorem (see, e.g. [Grimmett and Stirzaker \(1983\)](#)). The error in the estimated probability of failure will decrease with increasing number of sampled directions. Equation (A.31) can be used to determine the number of sampled directions required for an acceptably reliable estimate of the failure probability.

A.2.5.1 Determining the distance to the limit state

There are two search procedures in the Probabilistic Library to find λ_i , the distance to the limit state $Z = 0$ in the i^{th} sampled direction. The first method is illustrated in Figure A.8. Starting with $\lambda_i = 0$, the value of λ_i is incrementally increased by a fixed amount (the default is $\Delta\lambda = 1$). For the k^{th} iteration, the limit state function is evaluated at $\mathbf{U} = k\Delta\lambda \cdot \theta_i$. This continues until the limit state function becomes negative. For example, in Figure A.8, this occurs at $k = 6$. The algorithm then interpolates between $(k-1)\Delta\lambda$ and $k\Delta\lambda$ until it converges to $Z = 0$ with a convergence criteria on λ_i such that consecutive estimates lie within some small distance of each other (e.g. 0.001).

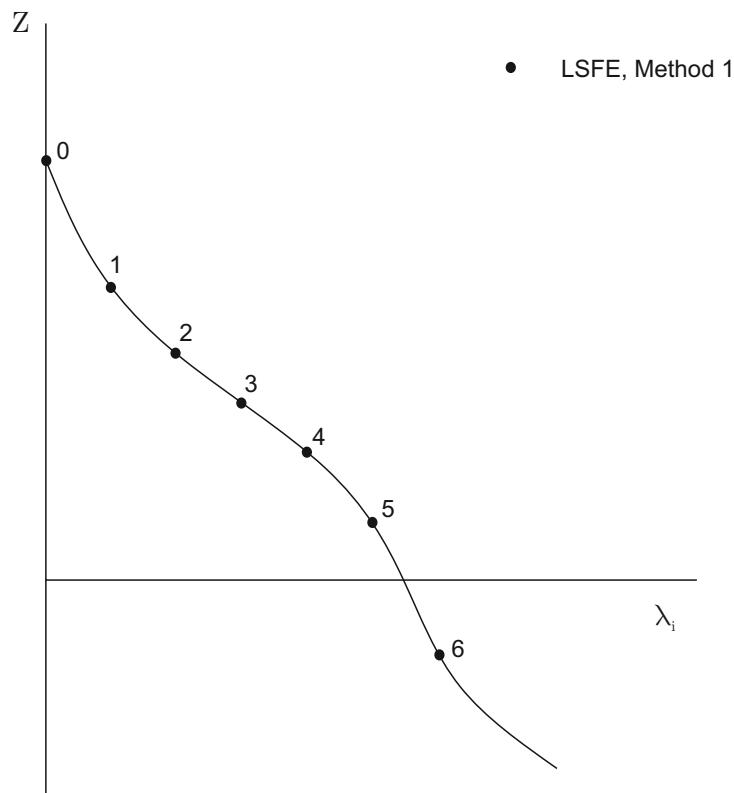


Figure A.8: Method 1 to determine λ_i , the distance to the limit state $Z = 0$ for the i^{th} sampled direction.

The second method is illustrated in Figure A.9. It is a gradient-based method that requires fewer limit state function evaluations (LSFEs) than Method 1 (described above). The algorithm first considers two values for λ_i : $\lambda_i^0 = 0$, and $\lambda_i^1 = \lambda_{step}$, where the default value is $\lambda_{step} = 3$. The superscript indicates the iteration in the algorithm. The limit state function is evaluated at both values, $Z^0 = Z(\lambda_i^0 \cdot \theta_i)$ and $Z^1 = Z(\lambda_i^1 \cdot \theta_i)$. Based on the two points $[\lambda_i^0, Z^0]$ and $[\lambda_i^1, Z^1]$, linear extrapolation is used to estimate the value λ_i^2 for which $Z = 0$. This extrapolation is indicated in Figure A.9 with the line ‘Gradient, step 1’. The limit state function is then evaluated again $Z^2 = Z(\lambda_i^2 \cdot \theta_i)$, and based on the points $[\lambda_i^1, Z^1]$ and $[\lambda_i^2, Z^2]$, linear extrapolation is again used to determine λ_i^3 , the value for which $Z = 0$ (in Figure A.9 this is the line ‘Gradient, step 2’). Note that for each iteration, the extrapolation is constrained so that if $\lambda_i^k - \lambda_i^{k-1} > \lambda_{step}$, then λ_i^k is set equal to $\lambda_i^{k-1} + \lambda_{step}$. The iterative process continues until either:

- ◊ $Z(\lambda_i^k \cdot \theta_i)$ becomes negative, at which point interpolation between $[\lambda_i^{k-1}, Z^{k-1}]$ and $[\lambda_i^k, Z^k]$ takes place in the same manner as Method 1,
- ◊ λ_i^{k-1} and λ_i^k differ by less than a small value ϵ (default is $\epsilon = 0.001$), which indicates one-sided convergence to $Z = 0$, or
- ◊ $\lambda_i^k > \lambda_{max}$ (default is $\lambda_{max} = 20$), at which point λ_i is set to λ_{max} , which is equivalent to a failure probability of zero.

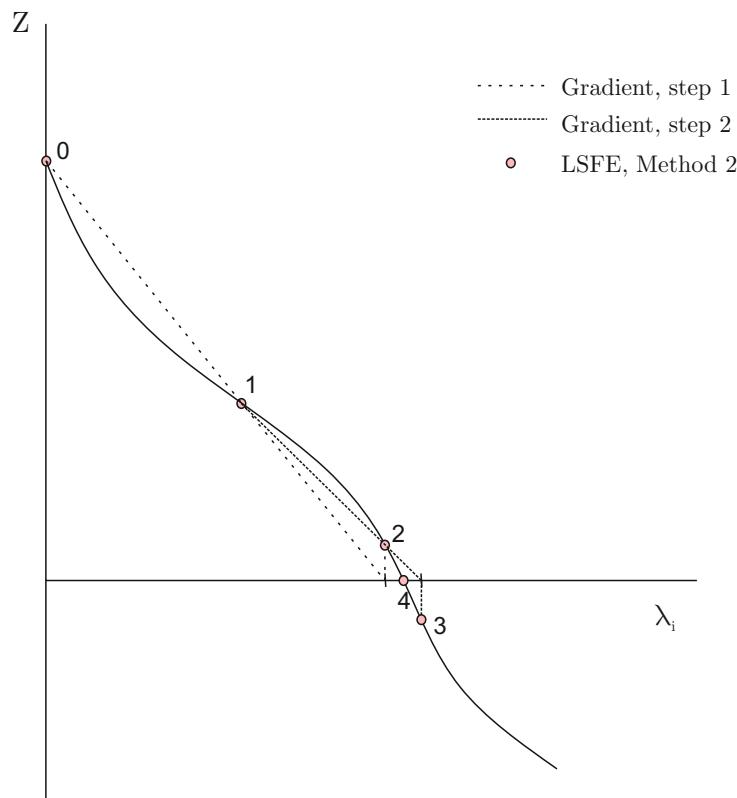


Figure A.9: Method 2 (gradient-based) to determine λ_i , the distance to the limit state $Z = 0$ for the i^{th} sampled direction.

The situation can arise where the gradient at the first iteration (the one derived using $[\lambda_i^0, Z^0]$ and $[\lambda_i^1, Z^1]$) is positive. This means that for increasing distance in the failure space, the limit state function is getting further away from $Z = 0$. If this occurs, the algorithm evaluates the limit state function at $\mathbf{U} = \lambda_{max} \cdot \theta_i$. If the value of the limit state function is positive, λ_i is set equal to λ_{max} . If the value is negative, we set $\lambda_i^2 = \lambda_i^1 + \lambda_{step}$, and the iterative process continues to find λ_i .

A.2.6 First-order reliability method (FORM)

The term first-order refers to the linearisation of the limit state function, as previously described in [section A.1](#). This linearisation takes place at a location referred to as the design point. A "location" in this case refers to a specific realization x_1, \dots, x_n of the X -variables, or u_1, \dots, u_n of the U -variables. The design point is the location along the limit state ($Z = 0$), where the probability density is maximal. This location is not known in advance and is determined via an iterative procedure, which will be explained in this section.

The FORM procedure is generally executed in the standard normal space (U -variables). The standard normally distributed variables have by definition a mean value of zero and a standard deviation of 1, and are mutually independent. The advantage of working in the standard normal space is that in this space the design point has a clear interpretation. Namely, in the standard normal space, the design point is the location along the limit state ($Z = 0$) which is closest to the origin (see [Figure A.10](#) for an illustration). This can be easily explained by the fact that for standard normally distributed variables the density is highest for $u = 0$ and decreases with increasing value of $|u|$. So in the u space the density decreases with increasing distance from the origin. Therefore, the design point is the point along the limit state that is closest to the origin.

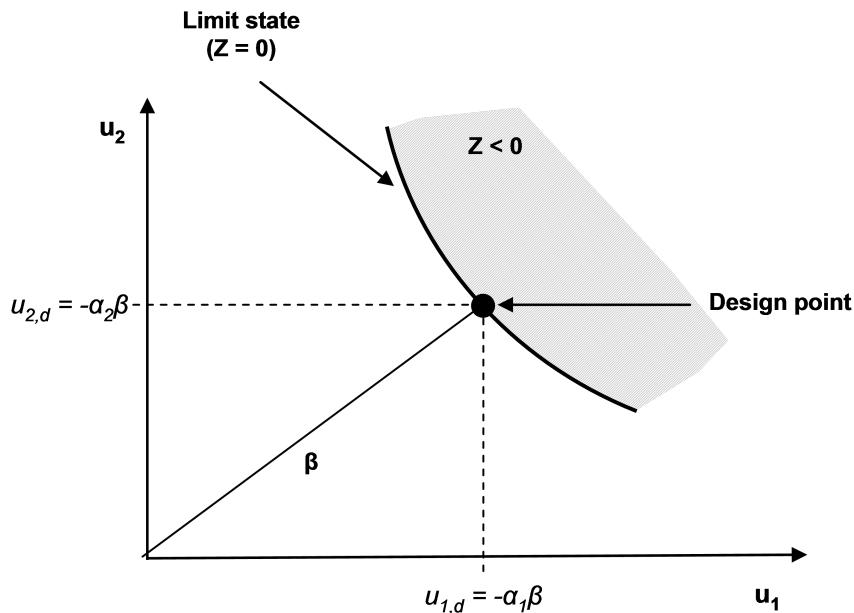


Figure A.10: Illustration of the design point in the U -space

The distance from the origin to the design point is equal to the reliability index β that was introduced in [section 5.2](#). This means if the location of the design point is known, the reliability index, β , is known and hence the estimated probability of failure can be derived from equation (5.8). This is how the probability of failure is estimated in the FORM method. Note that this is an estimate, and not the precise probability of failure. This is because the limit state function ($Z = 0$), was linearized to provide the estimate. The error in the estimate therefore depends on the extent in which the real limit state function is non-linear. This is illustrated in [Figure A.11](#) below, where the solid line indicates the true limit state and the dashed line represents the FORM approximation. The shaded area represents the true failure domain, the area to the upper right of the dashed line is the assumed failure domain.

The reason why FORM in general provides good estimates of the failure probability is the fact

that the linearisation is done in the design point, which means in the vicinity of the design point the linear Z -function is a good approximation of the real Z -function. The design point is the location on the limit state with the highest probability density. This means the failure events with the highest probability of occurrence will generally be in the vicinity of the design point. So, the linearised Z -function is generally a good approximation for the areas that matter the most in terms of probability of failure.

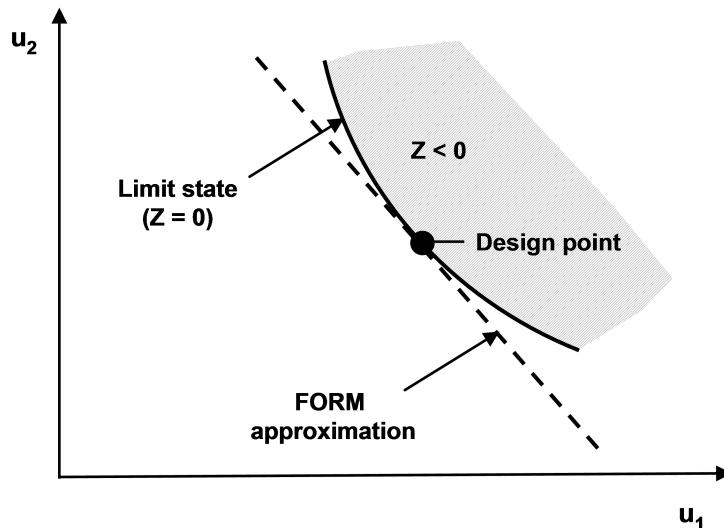


Figure A.11: Illustration of the FORM approximation

As demonstrated in [section A.1](#), the linearized limit state function Z_L is essentially a hyperplane and is described mathematically as follows:

$$Z_L = \beta + \alpha_1 U_1 + \dots + \alpha_n U_n = \beta + \sum_{i=1}^n \alpha_i U_i \quad (\text{A.32})$$

Note that the α -values have been normalized, as described in [section A.1](#). This means the sum of the squares of the α -values is equal to 1. In the remainder of this document, α -values are always assumed to be normalized in case they are used as coefficients in an equation with U -variables (unless mentioned otherwise). Given that the design point is the point along the $Z = 0$ line that lies closest to the origin, the coordinates of the design point can be determined using geometry as follows:

$$u_{d,i} = -\alpha_i \beta \quad ; i = 1 \dots n \quad (\text{A.33})$$

where $u_{d,i}$ is the value of the i^{th} random variable in the design point. As demonstrated in [section 5.2](#), the probability of failure can be estimated directly from the value of β :

$$P(Z < 0) = \Phi(-\beta) = 1 - \Phi(\beta) \quad (\text{A.34})$$

where Φ is the standard normal distribution function. In other words, once the design point is known, the Z -function can be linearised as described in [section A.1](#), and subsequently the probability of failure can be estimated from equation (A.34). The challenge in the FORM procedure is therefore not to compute the failure probability but to locate the design point.

The procedure to locate the design point is described below. The procedure and formulas will be presented in general form. Furthermore, the procedure will be clarified in a number of

figures for the following example Z -function:

$$Z = 5 - U_1^{0.8}U_2^{1.2}$$

where U_1 and U_2 are standard normal random variables. [Figure A.12](#) shows the contourlines of this Z -function. Generally, these contour lines are not known in advance, otherwise the search for the design point would be straightforward. Therefore, an iterative search procedure is required. The procedure starts at a "user defined" starting location in the U -space and jumps to a selected location in each following iteration step. In other words: in each iteration step the location in the U -space is determined that will serve as the starting point for the next iteration step. The procedure ends when the design point is found. Each iteration step consists of the following five sub-steps:

The five steps in a FORM iteration.

- 1 Linearisation of the Z -function in u^t , where u^t is the starting location of iteration t ;
- 2 Normalisation of the linearised Z -function in u^t ;
- 3 Estimation of the location of the design point, based on the Z -function of step 2;
- 4 Selection of location u^{t+1} , which will serve as the starting location of iteration $t + 1$;
- 5 Verification if the iteration procedure has converged.

These five steps are described in more detail below. Note that location u^t refers to a vector of u -values: $u^t = (u_1^t, \dots, u_n^t)$.

[1] The starting location of each iteration, t , is determined in the previous iteration, $t - 1$. The starting location in the first iteration step can either be selected "arbitrarily" or by more advanced methods in which the U -space is partially explored in advance of the FORM procedure. In the current example the starting location in the first iteration step is chosen to be $u_i = 1; i = 1 \dots n$ (red dot in [Figure A.12](#)).

In each iteration, first the Z -function is determined for the selected location at the beginning of the iteration, i.e. $Z(u_1, \dots, u_n)$ is quantified. Subsequently, the Z -function is linearised in the current location. For this purpose, the partial derivatives of Z to the individual U -variables are quantified. Generally, the Z -function is too complex to have an analytical expression of the partial derivatives, which means a numerical estimation technique is required. For this purpose, the Z -function is evaluated for small perturbations (Δu) of the u -values as shown in [Figure A.13](#). The partial derivatives can then be estimated as follows:

$$\frac{\partial Z}{\partial u_i}(u_1, \dots, u_n) \approx \frac{Z(u_1, \dots, u_i + \Delta u_i, \dots, u_n) - Z(u_1, \dots, u_n)}{\Delta u_i}; i = 1 \dots n \quad (\text{A.35})$$

Note that equation (A.35) describes a one-sided discretisation method. The Probabilistic Library actually uses a two-sided method, in which also a negative perturbation is applied on u_i :

$$\frac{\partial Z}{\partial u_i}(u_1, \dots, u_n) \approx \frac{Z(u_1, \dots, u_i + 0.5\Delta u_i, \dots, u_n) - Z(u_1, \dots, u_i - 0.5\Delta u_i, \dots, u_n)}{\Delta u_i}; i = 1 \dots n \quad (\text{A.36})$$

A two-sided method is generally more robust (because of often non-smooth limit state functions), but requires approximately twice as much computation time. The linearised Z -function is described by:

$$Z_L = B + A_1 U_1 + \dots + A_n U_n \quad (\text{A.37})$$

In which the A -values are the partial derivatives as derived in equation (A.35) and B is derived by substituting the known Z -value in the current location (u_1, \dots, u_n) :

$$B = Z(u_1, \dots, u_n) - A_1 u_1 - \dots - A_n u_n \quad (\text{A.38})$$

The linearised function is (temporarily) assumed to be valid for the entire U -space. This results in linear contour lines as shown in Figure A.14.



Note: In the Probabilistic Library, the following methods are available to determine the starting location of the u -vector in the first iteration step:

- ◊ All values of the u -vector are equal to 0.0.
- ◊ All values of the u -vector are equal to 1.0.
- ◊ The vector is defined by the user.
- ◊ Method ray search: the start vector u is found iteratively on a ray specified by the u_{Ray} -vector. In the u_{Ray} -vector, the load variables take value 1.0 and the strength variables take value 0.0.
- ◊ Method sphere search: the start vector u is found iteratively and the search takes place within a sphere with the radius of 10. The start vector is then defined as the u -vector (within the sphere) for which the smallest value of the Z -function is found.

[2] Subsequently, the linearised Z -function is normalized by dividing equation (A.37) by $\|A\|$, i.e. the norm of the A -vector (as earlier described in section A.1). The normalized linear Z -function is described as:

$$Z_L = \beta + \alpha_1 U_1 + \dots + \alpha_n U_n \quad (\text{A.39})$$

In which:

$$\beta = \frac{B}{\|A\|}; \quad \alpha_i = \frac{A_i}{\|A\|}, i = 1 \dots n; \quad \|A\| = \sqrt{\sum_{i=1}^n A_i^2} \quad (\text{A.40})$$

The normalization changes the contour lines of the linearised Z -function (compare Figure A.14 with Figure A.15). The orientation of the lines is still the same, but the distances between the contour lines have changed. The location of the contour line $Z = 0$, however, remains the same.

[3] From the linear contour lines it is easy to estimate the location of the design point. This is done by drawing the line through the origin that is perpendicular to the contour line $Z_L=0$ (see Figure A.16). In formula this means the estimated location of the design point is as described in equation (A.33). The values of α and β in equation (A.33) are set equal to the ones derived from equation (A.40).

[4] The estimated location of the design point can be chosen as the next location in the iteration procedure. Note, however, that this is most likely not the actual location of the design point, since it was derived from the linearised Z -function and not from the real Z -function. This is the reason why the design point will not be located straight away, i.e. a number of iteration steps are required. This is also the reason why in practical applications generally a relaxation parameter, R , is used in each iteration step:

$$u^{t+1} = R u_d^t + (1 - R) u^t \quad (\text{A.41})$$

In which:

t = iteration step

R = the relaxation parameter ($0 \leq R \leq 1$)

u^t = the selected location at the beginning iteration step t

u^{t+1} = the selected location at the beginning iteration step $t + 1$

u_d^t = the estimated location of the design point in iteration step t

The functionality of the relaxation parameter can be explained as follows: in each iteration step, the Z -function is linearised in location u^t . The linearised function, Z_L is the tangent of the actual Z -function at location u^t . In the vicinity of u^t , Z_L is generally a good approximation of Z . However, with increasing distance from u^t , differences between Z and Z_L may increase, as can be seen from e.g. Figure A.2. Since the estimated location of the design point, u_d^t , is based on Z_L , this estimate may be unreliable if the distance between u^t and u_d^t is large. This might even lead to non-convergence of the iteration procedure. It is therefore better to prevent that the distance between two subsequent iteration steps becomes too large, and for this reason the relaxation parameter is used. The relaxation parameter helps making the iterative procedure more robust.

Figure A.17 demonstrates the application of the relaxation parameter. It shows the location at the beginning of the iteration, u^t , (red dot), the estimated location, u_d^t , of the design point (yellow dot) and the location at the beginning of the next iteration, u^{t+1} (green dot). Location u^{t+1} is chosen somewhere on the line between the current location and the estimated location of the design point. For values of $R < 0.5$, u^{t+1} will be closer to u^t for values of $R > 0.5$, u^{t+1} will be closer to u_d^t .

[5] Figure A.18 shows the resulting iteration steps of the example problem. The iteration procedure continues until location u_d^t satisfies the following 2 criteria:

$$\frac{|Z(u_d^t)|}{\|A\|} = |Z_L(u_d^t)| < \varepsilon_1 \quad (\text{A.42})$$

$$\beta^t - \varepsilon_2 < \|u_d^t\| < \beta^t + \varepsilon_2 \quad (\text{A.43})$$

where:

$\varepsilon_{1,2}$ = Small numbers, quantifying convergence criteria

β^t = Estimate of reliability index β in iteration step 2

Criterion (A.42) guarantees that the Z -function is sufficiently close to 0, i.e. that u_d^t is on (or in the neighbourhood of) the limit state $Z = 0$. Note that for this purpose the value of Z is normalized by dividing it by the norm of the vector of A -values. The second criterion guarantees that the distance from u_d^t to the origin is (approximately) equal to the estimated reliability index β , which makes u_d^t the point on $Z = 0$ with the highest probability density.

Note: the FORM procedure has the advantage that it requires relatively little computation time, i.e. a relatively small number of Z -function evaluations. The disadvantage of this method is that the iterative algorithm sometimes does not converge and results may become less accurate. This is especially the case if the Z -function is highly non-linear.

Note: when FORM does not converge, then the results are corrected by averaging the results from the last ten iterations. The procedure is described below assuming n iterations and m random variables:



- ◊ For each variable determine the mean value of the last ten u -values: $u_k = \frac{1}{10} \sum_{i=n-9}^n u_{i,k}$ and $k = 1 \dots m$.
- ◊ Determine the mean value of the last ten reliability indices and determine the sign of the resulting value (the sign is -1.0 when the mean value is less than zero and the sign is 1.0 otherwise).
- ◊ Calculate the final reliability index as: $\beta = \text{sign} \cdot \sqrt{\sum_{k=1}^m u_k^2}$.
- ◊ For each variable calculate the final α -value as: $\alpha_k = -\frac{u_k}{\beta}$.

It has been shown that the above correction leads to more stable results.

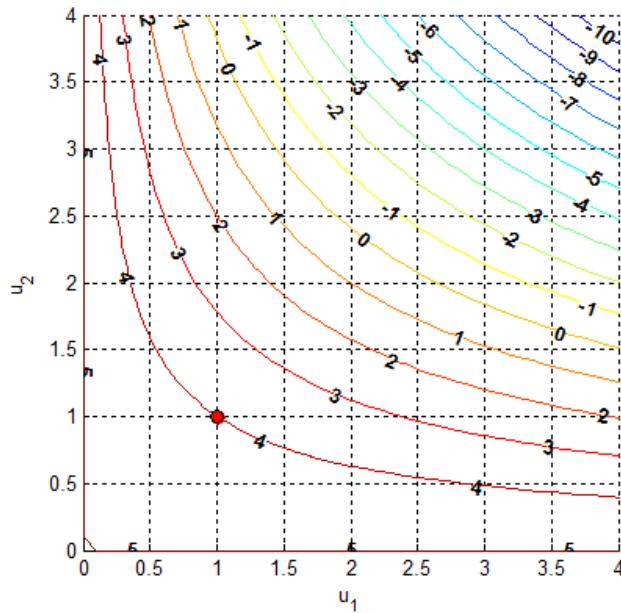


Figure A.12: Contour lines of the example Z-function and the starting location (red dot) of the FORM procedure

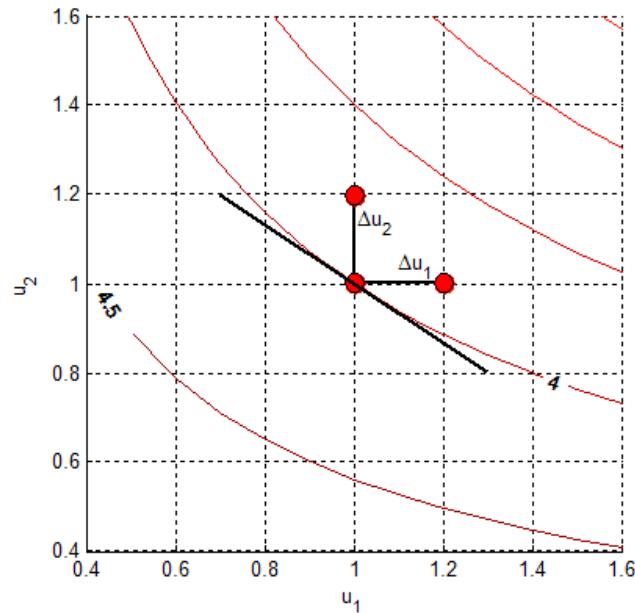


Figure A.13: Sampling the Z -function in all directions to estimate the derivative of Z to all u -variables.

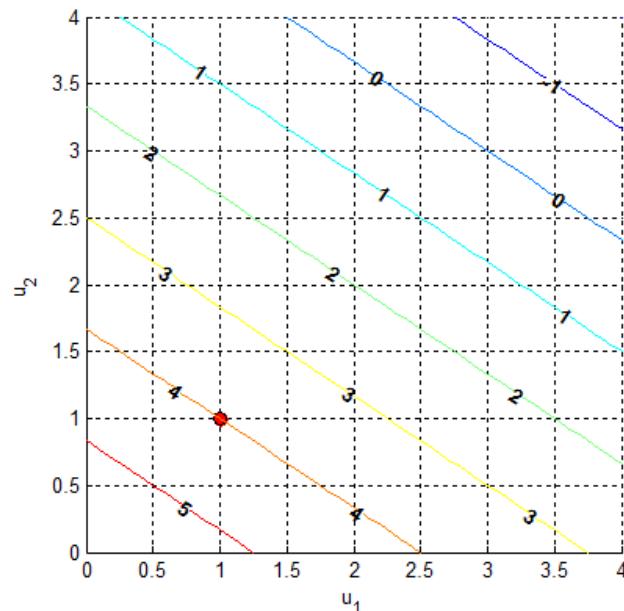


Figure A.14: Contour lines of the linearised Z -function

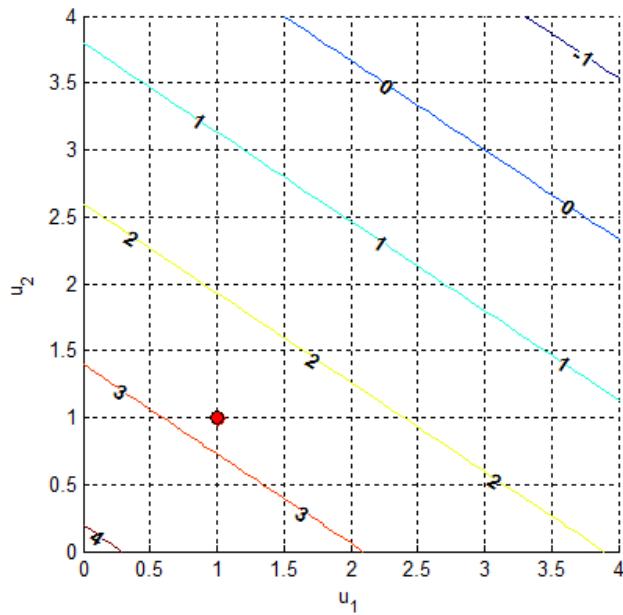


Figure A.15: Contour lines of the normalised linearised Z -function

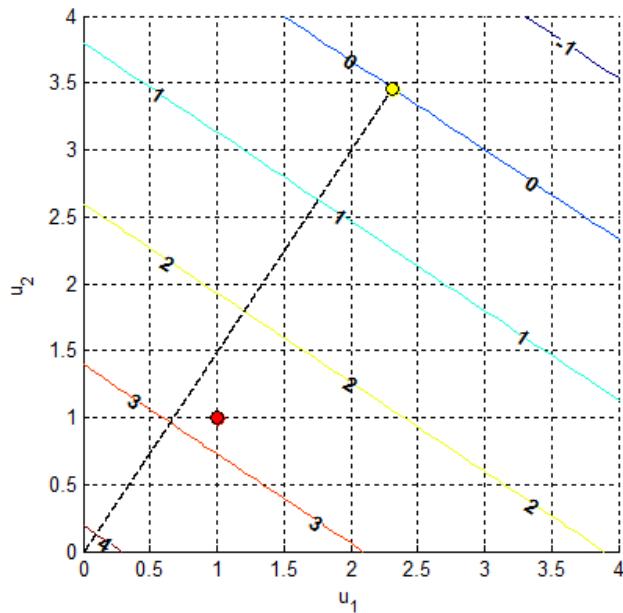


Figure A.16: Estimated location of the design point based on the normalised linearised Z -function

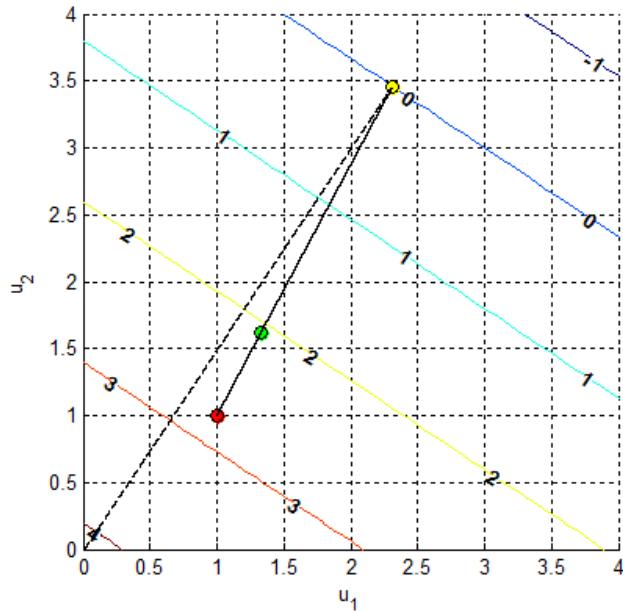


Figure A.17: Process of relaxation. The red dot is the location of the current iteration step, the yellow dot is the estimated location of the design point based on the normalised linearised Z-function. The green dot shows the selected location of the next iteration, which is somewhere on the line between the red dot and the yellow dot.

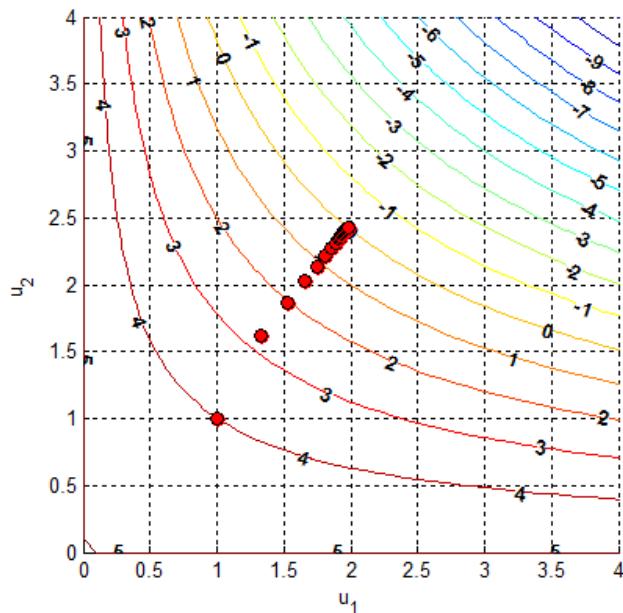


Figure A.18: Resulting steps in the iteration procedure.

A.2.7 Computing α values for other methods than FORM

In the previous section it was demonstrated that the FORM procedure not only provides an estimate of the probability of failure, but also a design point with a set of associated α -variables that provide information on the relative influence of the random variables on the reliability

index β (see also section A.1 on the meaning of α -variables). In section A.3 it will be demonstrated that these α -variables are very practical for estimating failure probabilities of systems that consist of a set of components.

The other probabilistic techniques that were described in the previous sections, i.e. numerical integration and the various Monte Carlo techniques, do not provide α -variables as output. Nevertheless, there are methods available to estimate α -variables for Monte Carlo methods and numerical integration (see, e.g., Van Gelder [2002]). For example, for Monte Carlo, the following methods can be applied:

- 1 Centre of gravity,
- 2 Method of angles and
- 3 Nearest to the mean.

The third method can also be used for numerical integration, as will be explained below. These methods all take into account the fact that the design point is the location in the failure domain that is closest to the origin (in the U -space). For all methods, the quality of the estimates increases with increasing number of samples. The methods are explained below:

Centre of gravity

Suppose a crude Monte Carlo run is done with n samples of which M lead to failure. This means there are M sampled combinations of u_1, \dots, u_n , for which $Z(u_1, \dots, u_n) < 0$. For each of the n random variables the mean value over the M failure-events is derived as follows:

$$\bar{u}_j = \frac{1}{M} \sum_{i=1}^M u_{ij} \quad ; j = 1 \dots n \quad (\text{A.44})$$

where u_{ij} is the i^{th} failure sample of the j^{th} random variable. The resulting point $\bar{u} = (\bar{u}_1, \dots, \bar{u}_n)$ is the centre of gravity in the failure domain in the U -space. This is an estimate of the probability weighted mean of the locations in the failure domain. Note that the actual probability weighted mean is equal to:

$$\bar{u}_j = \frac{1}{P[Z < 0]} \int_{Z<0} f_U(u) u_j du \quad (\text{A.45})$$

From the estimated centre of gravity this location a line is drawn towards the origin in the U -space (see Figure A.19). The location where this line crosses the limit state ($Z = 0$), is the estimated location of the design point. This guarantees that the first characteristic of the design point, i.e. that it is located on the limit state $Z = 0$, is taken care of. The second characteristic, that it is the location on the limit state with the highest density is not guaranteed. However, the use of the centre of gravity makes that the estimated location of the design point is likely to be close to the real design point. The likelihood increases with increasing number of samples.

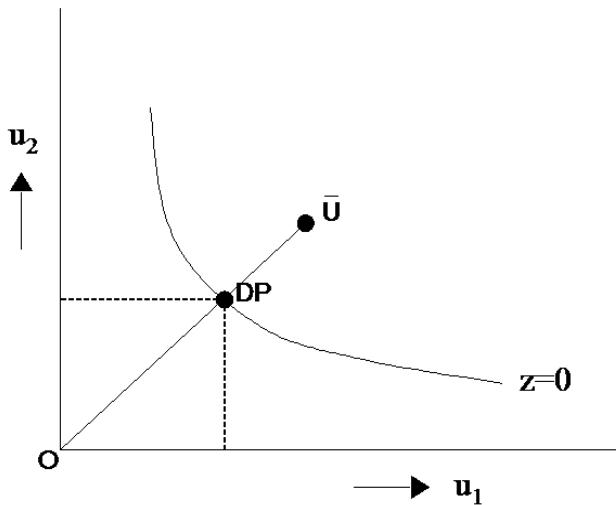


Figure A.19: Schematic view of the method centre of gravity.

If importance sampling has been applied in the sampling procedure, the method needs to be corrected for:

$$\bar{u}_j = \frac{1}{M} \sum_{i=1}^M u_{ij} \frac{f_U(u_{i1}, \dots, u_{in})}{h_U(u_{i1}, \dots, u_{in})} ; i = 1 \dots n \quad (\text{A.46})$$

where f_U is the probability density function of vector U and h_U is the applied density function of the importance sampling method.

Method of angles

The method of angles is similar to the method of centre of gravity. For each of the M samples that lead to failure the angle with the origin in the U -space is derived. After completion of the MC-procedure the mean angle of all M samples is derived. Note that the mean is derived with respect to the sine and cosine of the angles of all samples.

Method "nearest to the mean"

In the method of "nearest to the mean" the distance to the origin of all samples in the failure domain is derived:

$$|u_i| = \sqrt{\sum_{j=1}^n u_{ij}^2} ; i = 1 \dots M \quad (\text{A.47})$$

The sample with the smallest distance to the origin is taken to be the design point. This method can also be applied for other Monte Carlo techniques (directional sampling, importance sampling) and even for numerical integration. In the latter case, the design point is taken equal to the grid point in the failure domain that is closest to the origin in U -space.

A.2.8 Rationale

In the Probabilistic Library, a variety of probabilistic techniques has been implemented, including the first order reliability method (FORM), various Monte-Carlo techniques (crude, directional sampling, importance sampling) and numerical integration. Each of these techniques

requires a considerable number of evaluations of the Z -function at (randomly) selected x -values. The choice of the most suitable probabilistic computation technique depends on the problem under consideration.

If the computation time of one Z -function evaluation is significant, crude Monte Carlo and numerical integration are generally not the preferred candidates because both methods generally require a large number of Z -function evaluations. For crude Monte Carlo, the required number of Z -function evaluations is inversely proportional to the failure probability. This is because a small probability of failure means it takes a large number of samples to obtain even a single failure event and it takes more than one failure event to obtain a reliable estimate of the failure probability. For numerical integration, the number of Z -function evaluations is defined by the number of random variables and the number of grids for each random variable. Generally, numerical integration is too time-consuming if more than just a few random variables are involved. In theory, Monte Carlo and numerical integration are exact methods but in practice some error can be expected because the number of Z -function evaluations is limited.

Directional Sampling is a more advanced Monte Carlo method in comparison with crude Monte Carlo. For most practical problems it reduces the amount of Z -function evaluations in comparison with crude Monte Carlo. For a large number of random variables, the efficiency of directional sampling decreases (see e.g. [Waarts \(2000\)](#), pp 73). Importance sampling is another efficient Monte Carlo variant. The efficiency of importance sampling is accommodated by prescience of the location of the limit state function. Without prescience, the performance of importance sampling techniques is volatile.

FORM has the advantage that it requires relatively little computation time. The disadvantage of this method is that the iterative algorithm to find the design point sometimes does not converge or converges to a local design point. Furthermore, the Z -function is linearised in the method, which means errors are introduced if the actual Z -function is highly non-linear.

Combining two different probabilistic methods may result in the combined advantage of the underlying methods. For instance, a relatively fast method like FORM can be applied first to locate the design point and subsequently a more precise method like importance sampling can be applied to derive the probability of failure by sampling in the vicinity of the design point. FORM is usually more accurate in estimating design points, whereas a more robust Monte Carlo method will sometimes provide a more accurate estimate of the failure probability. Vice versa, Monte Carlo sampling can be applied to provide a starting point for FORM in the neighbourhood of the design point, to increase the chance that FORM converges to the correct design point. This increases the robustness of the FORM procedure.

A.3 Combining failure probabilities for components - generic methods

In this section, generic methods for the combination of the results for individual components are described. The general formulations of failure probabilities for parallel and series systems of k components are as follows:

$$\text{Series: } P_f = P [Z_1 < 0 \cup \dots \cup Z_k < 0] = P \left[\bigcup_{i=1}^k Z_i < 0 \right] = 1 - P \left[\bigcap_{i=1}^k Z_i \geq 0 \right] \quad (\text{A.48})$$

$$\text{Parallel: } P_f = P [Z_1 < 0 \cap \dots \cap Z_k < 0] = P \left[\bigcap_{i=1}^k Z_i < 0 \right] = 1 - P \left[\bigcup_{i=1}^k Z_i \geq 0 \right] \quad (\text{A.49})$$

If the events $[Z_i < 0]$, $i = 1 \dots k$ are mutually independent, this can be simplified to:

$$\text{Series: } P_f = 1 - \prod_{i=1}^k \{1 - P[Z_i < 0]\} \quad (\text{A.50})$$

$$\text{Parallel: } P_f = \prod_{i=1}^k P[Z_i < 0] \quad (\text{A.51})$$

The failure probabilities, $P[Z_i < 0]$, for the individual components are determined by the probabilistic computation techniques as described in section A.2. System analysis for mutually independent components is therefore a relatively straightforward procedure. If the components are mutually correlated, the complexity of the system analysis increases. The correlations need to be taken into account as it increases the probability of failure of parallel systems and decreases the probability of failure of series systems. The following sections describe various general techniques that can be applied to carry out system analysis for systems with mutually correlated components.

A.3.1 Combining n system components: the Hohenbichler method

In this section, the Hohenbichler method for combining multiple components is highlighted.

A.3.1.1 Introduction

The Hohenbichler method initially is a method for computing conditional probabilities of two Z -functions: $P(Z_2 < 0 | Z_1 < 0)$, taking into account the mutual correlation between these two Z -functions. The application of this method can be extended to compute failure probabilities of:

- 1 A parallel system of two components;
- 2 A series system of two components;
- 3 Parallel and series systems of multiple components.

This is explained as follows:

[1] A parallel system with two components refers to a system in which both components must fail in order for failure to occur (keyword: AND). That is, the probability of failure is given as follows:

$$P(F) = P(Z_1 < 0 \cap Z_2 < 0) \quad (\text{A.52})$$

A parallel system and the schematization of the associated failure probability $P(Z_1 < 0 \cap Z_2 < 0)$ is schematically depicted in Figure A.20 below.

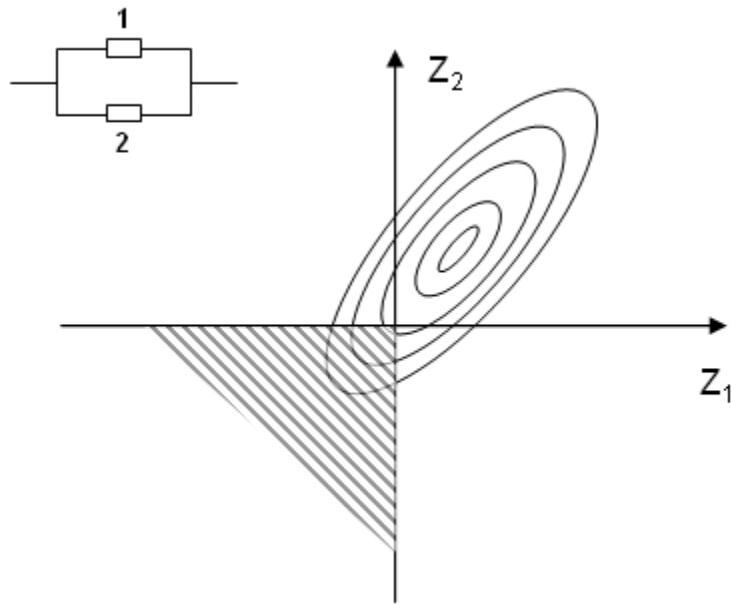


Figure A.20: Failure domain for a parallel system of two components – the shaded area indicates the combinations of Z -values that will lead to failure of the system.

The failure probability of this system can be written equivalently as the product of a probability and a conditional probability:

$$P(F) = P(Z_1 < 0) \cdot P(Z_2 < 0 | Z_1 < 0) \quad (\text{A.53})$$

The first term, $P(Z_1 < 0)$ can be computed with the methods as described in section A.2. The second term, $P(Z_2 < 0 | Z_1 < 0)$, can be determined with the Hohenbichler method, as will be demonstrated in subsequent sections. This shows that the Hohenbichler method can also be applied to compute the failure probability of a parallel system of two components.

[2] A series system with two components refers to a system in which at least one component must fail in order for failure to occur (keyword: OR):

$$P(F) = P(Z_1 < 0 \cup Z_2 < 0) \quad (\text{A.54})$$

This probability can be rewritten as follows:

$$P(F) = P(Z_1 < 0 \cup Z_2 < 0) = P(Z_1 < 0) + P(Z_2 < 0) - P(Z_1 < 0 \cap Z_2 < 0) \quad (\text{A.55})$$

The first two terms on the right hand side of equation (A.55) describe failure probabilities of single components, which can be derived with the techniques that were described in section A.2. The last term describes a parallel system of two components, for which the computational method was described in [1]. This shows that the Hohenbichler method can also be applied to compute the failure probability of a series system of two components.

[3] Consider a series system of n components. The failure probability for the system is given by:

$$P(F) = P(Z_1 < 0 \cup Z_2 < 0 \cup Z_3 < 0 \cup \dots \cup Z_n < 0) \quad (\text{A.56})$$

If we define $Z_{12} = Z_1 \cup Z_2$, this equation can be rewritten as an arbitrary system of $n - 1$ components:

$$P(F) = P(Z_{12} < 0 \cup Z_3 < 0 \cup \dots \cup Z_n < 0) \quad (\text{A.57})$$

Repeating this procedure $n - 1$ times will result in a system of one component. So, the probability of failure for a system of n components can be derived by the successive combining of combinations of two components. The method of Hohenbichler can be used to combine two components, as demonstrated above, and therefore it can also be used to combine n components of a series system. A similar approach can be applied for a parallel system of n components. In other words: successive application of the method can than be used to compute the probability of failure of a system of n components.

The basic principles of the Hohenbichler method for computing the conditional probability of failure of two components is described in [section A.3.1.2](#). The follow-up sections elaborate on some of the finer details.

Note: the Hohenbichler method makes use of linearisation of the Z -functions, as described in [section A.1](#). The probability of failure for a system as derived by the method of Hohenbichler is therefore an approximation of the real probability of failure. Errors made in the approximation will depend on the system under consideration.



A.3.1.2 Hohenbichler method for a system with fully correlated variables

As stated in the previous section, the Hohenbichler method initially is a method for computing conditional probabilities of two Z -functions: $P(Z_2 < 0 | Z_1 < 0)$, taking into account the mutual correlation, ρ , between these two Z -functions. In this section, first the computational procedure will be described, followed by a detailed explanation.

The two Z -functions are described by a set of U -variables. In this section, the set of U -variables are assumed to be the same for the two Z -functions. In [section A.3.1.4](#) the slightly more complex case will be dealt with in which the set of U -variables are different for the two Z -functions.

Computational procedure

The following information is available from the single component probabilistic analysis as described in [section A.2](#):

- ◊ The reliability index β_1 of Z_1
- ◊ The reliability index β_2 of Z_2
- ◊ The influence variable, α , for each random variable involved.

First, the correlation between the Z -functions of the two components needs to be quantified:

$$\rho(Z_1, Z_2) = \sum_{j=1}^n \alpha_{1j} \cdot \alpha_{2j} \quad (\text{A.58})$$

Subsequently, the conditional probability is computed by solving the following integral:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{\int_{\beta_1}^{\infty} \Phi\left(-\frac{\beta_2 - \rho u_1}{\sqrt{1-\rho^2}}\right) \phi(u_1) du_1}{\Phi(-\beta_1)} \quad (\text{A.59})$$

where ϕ is the standard normal density function. The outcome of this integral can be approximated very accurately through numerical integration at relatively low computational costs.

Detailed explanation

Since the α - and β -values of the two Z -functions are known, the Z -functions can be described in the standard linearised form (see section A.1):

$$\begin{aligned} Z_1 &= \beta_1 + \alpha_{11}U_{11} + \dots + \alpha_{1n}U_{1n} \\ Z_2 &= \beta_2 + \alpha_{21}U_{21} + \dots + \alpha_{2n}U_{2n} \end{aligned} \quad (\text{A.60})$$

where U_{ij} refers to the j^{th} random variable of the i^{th} Z -function. The U -variables can be different for the different Z -functions. However, for the sake of simplicity, we assume for the moment that they are the same:

$$U_{1k} = U_{2k} \quad ; k = 1 \dots n \quad (\text{A.61})$$

In section A.3.1.4 the slightly more complex case will be dealt with in which $U_{1k} \neq U_{2k}$. Since U_{1j} and U_{1k} , $j \neq k$, are mutually independent, it can easily be verified that the correlation between Z_1 and Z_2 is equal to:

$$\rho(Z_1, Z_2) = \sum_{j=1}^n \alpha_{1j} \cdot \alpha_{2j} \quad (\text{A.62})$$

The linearised Z -functions can be re-written (see section A.1) as follows:

$$\begin{aligned} Z_1 &= \beta_1 - U_1 \quad ; U_1 = -(\alpha_{11}U_{11} + \dots + \alpha_{1n}U_{1n}) \\ Z_2 &= \beta_2 - U_2 \quad ; U_2 = -(\alpha_{21}U_{21} + \dots + \alpha_{2n}U_{2n}) \end{aligned} \quad (\text{A.63})$$

where U_1 and U_2 are two newly defined standard normally distributed variables. Because the β -values in equation (A.63) are constant, the correlation between the components Z_1 and Z_2 is equivalent to the correlation between the variables U_1 and U_2 :

$$\rho(Z_1, Z_2) = \rho(U_1, U_2) = \rho \quad (\text{A.64})$$

In other words, equation (A.63) is only valid if U_1 and U_2 are mutually correlated with correlation coefficient ρ . To assure that this is the case, U_2 is written as a function of U_1 :

$$U_2 = \rho U_1 + U_2^* \sqrt{1 - \rho^2} \quad (\text{A.65})$$

In this equation, U_2^* is also standard normally distributed and independent of U_1 . The first term in this equation represents the dependent part of U_2 and the second term represents the independent part. Note in equation (A.65) that if $\rho = 1$, $U_2 = U_1$ (100% correlated), and if $\rho = 0$, then $U_2 = U_2^*$ (100% uncorrelated). To verify the applicability of equation (A.65) it needs to be shown that [1] U_2 is standard normally distributed and [2] that U_1 and U_2 have a mutual correlation coefficient that is equal to ρ . To prove [1], we apply the following general rule (see, e.g. Grimmett and Stirzaker (1983)): If X and Y are independent normally distributed random variables, then $aX + bY$ is also normally distributed with a mean, μ , and standard deviation, σ , equal to:

$$\begin{aligned} \mu &= a\mu_X + b\mu_Y \\ \sigma &= \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2} \end{aligned} \quad (\text{A.66})$$

Application of this rule on equation (A.65), where U_1 and U_2^* are both normally distributed with mean 0 and standard deviation 1, gives:

$$\begin{aligned}\mu(U_2) &= \rho \cdot 0 + \sqrt{1 - \rho^2} \cdot 0 = 0 \\ \sigma(U_2) &= \sqrt{\rho^2 \cdot 1 + (1 - \rho^2) \cdot 1} = 1\end{aligned}\quad (\text{A.67})$$

Which proves that U_2 is standard normally distributed. To prove [2], the correlation coefficient between U_1 and U_2 is derived. The correlation coefficient of U_1 and U_2 is defined as:

$$\rho(U_1, U_2) = \frac{\text{cov}(U_1, U_2)}{[\sigma(U_1)\sigma(U_2)]} = \frac{\text{cov}(U_1, U_2)}{[1 \cdot 1]} = \text{cov}(U_1, U_2) \quad (\text{A.68})$$

The covariance of U_1 and U_2 is equal to:

$$\begin{aligned}\text{cov}(U_1, U_2) &= E[U_1 U_2 - \mu(U_1)\mu(U_2)] = E[U_1 U_2] \\ &= E\left[U_1 \left(\rho U_1 + U_2^* \sqrt{1 - \rho^2}\right)\right] \\ &= E\left[\rho U_1^2 + U_1 U_2^* \sqrt{1 - \rho^2}\right] = E[\rho U_1^2] \\ &= \rho E[U_1^2] = \rho \text{Var}[U_1] = \rho\end{aligned}\quad (\text{A.69})$$

Which proves that the application of equation (A.65) preserves the correlation between U_1 and U_2 and hence the correlation between Z_1 and Z_2 . The combination of equations (A.63) and (A.65) provides the following alternative description for function Z_2 :

$$Z_2 = \beta_2 - \rho U_1 - U_2^* \sqrt{1 - \rho^2} \quad (\text{A.70})$$

This expression represents a line in the $Z_2 = 0$ plane. The hatched area in Figure A.21 indicates the area in the U -space that contributes to the failure probability

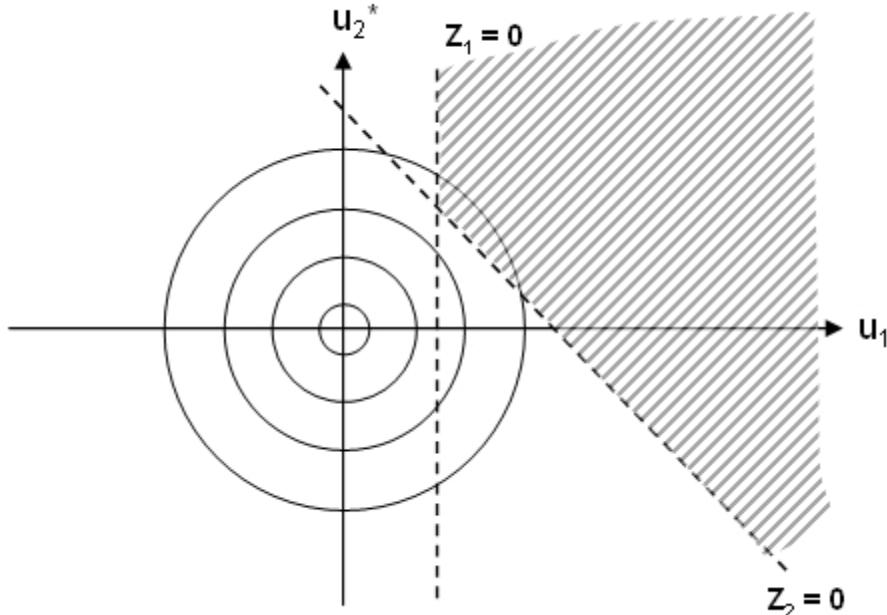


Figure A.21: The $Z_1 = 0$ and $Z_2 = 0$ contours in the U -space; hatched area indicates the area that contributes to the failure probability.

Consider the following general formulation for conditional failure probability:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{P[Z_2 < 0 \cap Z_1 < 0]}{P[Z_1 < 0]} \quad (\text{A.71})$$

From equation (A.63) it can be seen that $Z_1 < 0$ if and only if $U_1 > \beta_1$. This means:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{P[Z_2 < 0 \cap U_1 > \beta_1]}{P[U_1 > \beta_1]} = \frac{P[Z_2 < 0 \cap U_1 > \beta_1]}{\Phi(-\beta_1)} \quad (\text{A.72})$$

Substitution of equation (A.70) in equation (A.72) gives:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{P\left[\beta_2 - \rho U_1 - \sqrt{1 - \rho^2} U_2^* < 0 \cap U_1 > \beta_1\right]}{\Phi(-\beta_1)} \quad (\text{A.73})$$

The numerator can be computed by integration over all potential realisations of $U_1 > \beta$, using the theorem of total probability:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{\int_{\beta_1}^{\infty} P\left[\beta_2 - \rho u_1 - \sqrt{1 - \rho^2} U_2^* < 0\right] \phi(u_1) du_1}{\Phi(-\beta_1)} \quad (\text{A.74})$$

The probability in the numerator can be rewritten as follows:

$$\begin{aligned} P\left[\beta_2 - \rho u_1 - \sqrt{1 - \rho^2} U_2^* < 0\right] &= P\left[U_2^* > \frac{\beta_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right] \\ &= P\left[U_2^* < -\frac{\beta_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right] = \Phi\left(-\frac{\beta_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) \end{aligned} \quad (\text{A.75})$$

Substitution of equation (A.75) in equation (A.74) gives:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{\int_{\beta_1}^{\infty} \Phi\left(-\frac{\beta_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) \phi(u_1) du_1}{\Phi(-\beta_1)} \quad (\text{A.76})$$

Note that due to the symmetry of the normal density function, ϕ , this equation can be re-written as follows:

$$P[Z_2 < 0 | Z_1 < 0] = \frac{\int_{-\infty}^{-\beta_1} \Phi\left(-\frac{\beta_2 + \rho u_1}{\sqrt{1 - \rho^2}}\right) \phi(u_1) du_1}{\Phi(-\beta_1)} \quad (\text{A.77})$$

Both equations (A.76) and (A.77) can be solved by numerical integration at a relatively low computational cost. Alternative methods using FORM or Monte Carlo can be applied as well, but for this specific application numerical integration is recommended as it takes little computation time and it is very robust and reliable.

A.3.1.3 Derivation of equivalent α -values

The previous section describes the Hohenbichler method for combining the probability of failure of two single components. The result is the combined probability of failure. The goal of a systems approach to failure probability is to combine the failure probabilities of all the contributing components to determine the failure probability of the whole system. This combining of probabilities takes place in a sequential fashion. That is, first two components are combined into one component, and this new component is then combined with an additional component, and so on, until only one component (the entire system) remains.

The combining of failure probabilities over components relies on α -values, see equation (A.62). This means we require α -values for $Z_1 \cup Z_2$, in order to be able to quantify the correlation with a third component, Z_3 . The required α -values are referred to as 'equivalent' α -values. Basically, a Z -function description, Z^e , is required that represents the combined components Z_1 and Z_2 . This Z -function needs to have the same probability of failure as Z_1 and Z_2 :

$$P(Z^e < 0) = P(Z_1 < 0 \cup Z_2 < 0) \quad (\text{A.78})$$

The equivalent α -values should be such that they describe this Z -function in the standard linearised way:

$$Z^e = \beta^e + \alpha_1^e U_1 + \dots + \alpha_n^e U_n \quad (\text{A.79})$$

Computational procedure

To compute the equivalent α -values, the following procedure is applied:

[1] Derive conditional failure probability $P(Z_2 < 0 | Z_1 < 0)$ with the Hohenbichler method as described in the previous section using the following values as input: reliability index β_1 of Z_1 , reliability index β_2 of Z_2 and the influence variable, α , for each random variable involved. Subsequently, compute $P[Z_1 \cup Z_2]$ (or $P[Z_1 \cap Z_2]$ if a parallel system is considered) with the techniques described in section A.3. Also compute the associated "equivalent" reliability index, β^e :

$$\beta^e = \Phi^{-1}(P[Z_1 \cup Z_2]) \quad (\text{A.80})$$

[2] For each variable U_k , $k = 1 \dots n$, carry out steps [2a] and [2b]

[2a] Repeat the procedure of step 1 with input reliability indices $\beta_1' = \beta_1 + \varepsilon \alpha_{1k}$ and $\beta_2' = \beta_2 + \varepsilon \alpha_{2k}$, where ε is a small perturbation. Refer to the resulting reliability index as: $\beta_k^e(\varepsilon)$.

[2b] Compute the equivalent α -value of variable U_k as follows:

$$\alpha_k^e = \frac{\beta_k^e(\varepsilon) - \beta^e}{\varepsilon} \quad (\text{A.81})$$

The result is set of n equivalent α -values, $\alpha_1^e, \dots, \alpha_n^e$ (step 2) and an equivalent reliability index β^e (step 1). These values define the equivalent Z -function as described in equation (A.79).

Detailed explanation

The equivalent value β^e is the reliability index that is derived with the Hohenbichler method as described in section A.3.1.2. In order to derive the α -values of function Z^e , recall from section A.1 that the α -values of a Z -function are related to the reliability index β as follows:

$$\frac{\partial \beta^e}{\partial \bar{u}_k} = \alpha_k^e \quad (\text{A.82})$$

where \bar{u}_k is the mean value of variable U_k . In order to estimate the sensitivity of the equivalent reliability index β^e to small perturbations in the mean value of variable U_k , we require a set of definitions. First, consider the two following linear Z -functions of n components:

$$Z_i = \beta_i + \alpha_{i1}U_1 + \dots + \alpha_{ik}U_k + \dots + \alpha_{in}U_n \quad ; i = 1, 2 \quad (\text{A.83})$$

Furthermore, define $Z_{ik}(\varepsilon)$, $i = 1, 2$, as the Z -functions that result from increasing the mean value of variable U_k in equation (A.83) with a small perturbation ε . To analyse $Z_{ik}(\varepsilon)$, define:

$$U'_k = U_k + \varepsilon \quad (\text{A.84})$$

Since U_k is standard normally distributed, U'_k is normally distributed with mean ε and standard deviation 1. Function $Z_{ik}(\varepsilon)$ is obtained by replacing U_k in equation (A.83) by U'_k :

$$\begin{aligned} Z_{ik}(\varepsilon) &= \beta_i + \alpha_{i1}U_1 + \dots + \alpha_{ik}U'_k + \dots + \alpha_{in}U_n \quad ; i = 1, 2 \\ &= \beta_i + \alpha_{i1}U_1 + \dots + \alpha_{ik}(U_k + \varepsilon) + \dots + \alpha_{in}U_n \quad ; i = 1, 2 \\ &= (\beta_i + \alpha_{ik}\varepsilon) + \alpha_{i1}U_1 + \dots + \alpha_{ik}U_k + \dots + \alpha_{in}U_n \quad ; i = 1, 2 \end{aligned} \quad (\text{A.85})$$

The equivalent reliability index, β_k^e , for this perturbed system of two components is computed as follows:

$$\beta_k^e(\varepsilon) = \Phi^{-1}(1 - P[Z_{1k}(\varepsilon) < 0 \cup Z_{2k}(\varepsilon) < 0]) \quad (\text{A.86})$$

The equivalent α -value, α_k^e , can be derived from:

$$\alpha_k^e = \frac{\partial \beta_k^e}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon} [\Phi^{-1}(1 - P[Z_{1k}(\varepsilon) < 0 \cup Z_{2k}(\varepsilon) < 0])] \quad (\text{A.87})$$

Generally, Z -functions are too complex to derive equation (A.87) analytically. Therefore, a numerical approach is required in which the mean of variable U_i is perturbed by a small value ε_i and subsequently the change in the value of β^e is quantified.

From equations (A.83) and (A.85) it follows:

$$Z_{ik}(\varepsilon) = Z_i + \alpha_{ik}\varepsilon; i = 1, 2 \quad (\text{A.88})$$

So if Z_i has a reliability index β_i , $Z_{ik}(\varepsilon)$ has a reliability index equal to $\beta_i + \varepsilon\alpha_{ik}$. So Z_i and $Z_{ik}(\varepsilon)$ have different reliability indices, but the α -values of these two Z -functions are the same. This explains why in step 2 of the procedure as described in the previous section, Z -functions with reliability indices $\beta_i + \varepsilon\alpha_{ik}$ are evaluated with the Hohenbichler method to quantify the sensitivity of the equivalent Z -function to small perturbations ε in the mean value of variable U_k .

Example

Figure A.22 shows an example of two Z -functions, Z_1 and Z_2 , and the equivalent Z -function, Z^e . The variables of the Z -functions are displayed in Table A.3. The failure probability $P(Z_1 < 0 \cup Z_2 < 0)$ as derived with the Hohenbichler method in this example is equal to $6.120 \cdot 10^{-3}$, whereas the failure probability based on Monte Carlo sampling with 10^8 samples (i.e. a very accurate method) is equal to $6.121 \cdot 10^{-3}$. This shows that the Hohenbichler method in principle is an exact method for combining 2 linear Z -functions (linear as a function of the u -variables). However, errors will be introduced [a] if the real Z -functions are non-linear or [b] when combining more than 2 components.

Table A.3: Description of the Z -functions of Figure A.22.

Function	Before normalization			After normalisation		
Variable:	α_1	α_2	β	α_1	α_2	β
Z_1	-2	-1	6	-0.89	-0.45	2.68
Z_2	-1	-2	6	-0.45	-0.89	2.68

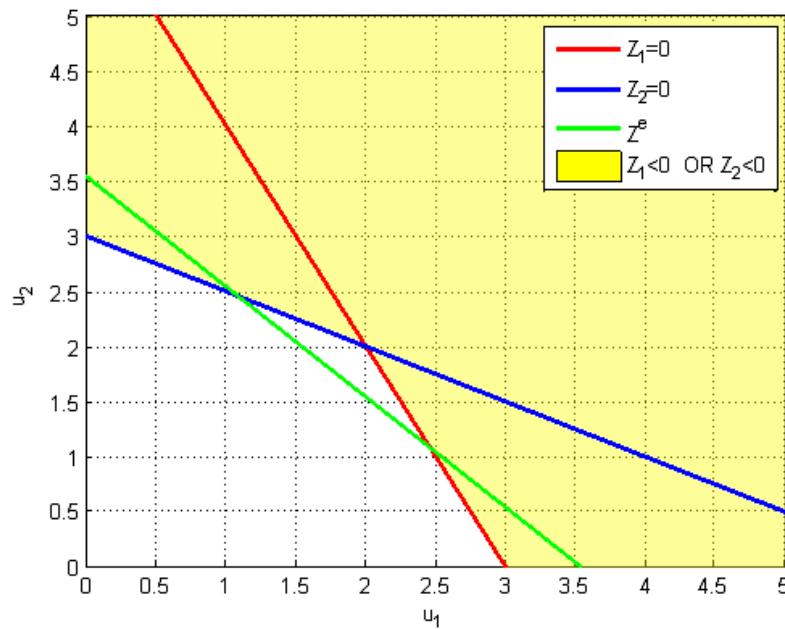


Figure A.22: Replacement of " $Z_1 < 0 \cup Z_2 < 0$ " with " $Z^e < 0$ ". Failure domains of the three Z -functions are on the upper right side of the respective lines.

A.3.1.4 Hohenbichler method for the general case of partial correlation

In the previous sections, Z_1 and Z_2 were functions of the same variables U_1, \dots, U_n . In the current section the more general case is considered:

$$\begin{aligned} Z_1 &= \beta_1 + \alpha_{11}U_{11} + \dots + \alpha_{1n}U_{1n} \\ Z_2 &= \beta_2 + \alpha_{21}U_{21} + \dots + \alpha_{2n}U_{2n} \end{aligned} \tag{A.89}$$

where U_{ik} refers to the k^{th} random variable of the i^{th} Z -function. In this case, variables U_{1k} and U_{2k} are partially correlated. All other correlations are equal to zero:

$$\rho(U_{1j}, U_{2k}) = \begin{cases} \rho_{12k} & ; j = k \\ 0 & ; j \neq k \end{cases} \quad (\text{A.90})$$

For components 1 and 2, the k^{th} random variable (i.e. U_{1k} and U_{2k}) in principle refers to the same load or strength variable, but the sampled values can be different because they refer to different components. For instance, the k^{th} variable may refer to the thickness of a clay layer. This thickness will be different for different dike segments.

Note that equation (A.89) also covers the case in which the two Z -functions depend on different sets of random variables, for instance because two dike segments are combined that are situated along different water systems, or if one component is a dike segment and the other one is a dune segment. In that case, the combined sets of random variables will be used in equation (A.89) and some of the α -values will be equal to zero.

Computational procedure

To compute the equivalent α -values, the following procedure is applied:

[1] Derive conditional failure probability $P(Z_2 < 0 | Z_1 < 0)$ with the Hohenbichler method as described [section A.3.1.2](#) using the following values as input: reliability index β_1 of Z_1 , reliability index β_2 of Z_2 and the influence variable, α , for each random variable involved. Subsequently, compute $P[Z_1 \cup Z_2]$ (or $P[Z_1 \cap Z_2]$ if a parallel system is considered) with the techniques described in [section A.3](#). Also compute the associated "equivalent" reliability index, β^e :

$$\beta^e = \Phi^{-1}(P[Z_1 \cup Z_2]) \quad (\text{A.91})$$

[2] For each variable U_k , $k = 1 \dots n$, carry out steps [2a] – [2e]:

[2a] Repeat the procedure of step 1 with input reliability indices $\beta'_1 = \beta_1 + \varepsilon\alpha_{1k}$ and $\beta'_2 = \beta_2 + \varepsilon\rho_{12k}\alpha_{2k}$. The resulting equivalent reliability index is referred to as: β_c .

[2b] Compute the following equivalent α -value:

$$\alpha_k^I = \frac{\beta_c - \beta^e}{\varepsilon} \quad (\text{A.92})$$

[2c] Repeat the procedure of step 1 with input reliability indices $\beta'_1 = \beta_1$ and $\beta'_2 = \beta_2 + c$, with c equal to

$$c = \varepsilon\alpha_{2,k} \sqrt{1 - (\rho_{12k})^2} \quad (\text{A.93})$$

The resulting equivalent reliability index is referred to as: β^u .

[2d] Compute the following equivalent α -value:

$$\alpha_k^{II} = \frac{\beta^u - \beta^e}{\varepsilon} \quad (\text{A.94})$$

[2e] Compute:

$$\alpha_k^e = \sqrt{(\alpha_k^I)^2 + (\alpha_k^{II})^2} \quad (\text{A.95})$$

The result of steps 2a – 2e is set of n equivalent α -values, $\alpha_1^e \dots \alpha_n^e$.

[3] Carry out a normalization step

$$\alpha_{k;final}^e = \frac{\alpha_k^e}{\sqrt{\sum_{j=1}^n (\alpha_j^e)^2}} ; k = 1 \dots n \quad (\text{A.96})$$

Detailed explanation

From equation (A.90) it follows that the correlation between Z_1 and Z_2 is equal to:

$$\rho(Z_1, Z_2) = \sum_{k=1}^n \alpha_{1k} \alpha_{2k} \rho_{12k} \quad (\text{A.97})$$

In order to determine the probability of failure $P(Z_1 < 0 \cup Z_2 < 0)$, the exact same method as described in section A.3.1.2 is used. In order to derive the equivalent α -values, the procedure becomes somewhat more complicated than the procedure with full correlation that was described in section A.3.1.3. The approach is to describe the U -variables of Z_2 as a function of the U -variables of Z_1 :

$$u_{2k} = U_{1k} \rho_{12k} + U_{2k}^* \sqrt{(1 - \rho_{12k}^2)} ; k = 1 \dots n \quad (\text{A.98})$$

The first term in this function represents the part of U_{2k} that is fully correlated to U_{1k} , the second term describes the part that is fully uncorrelated to U_{1k} . Variable U_{2k}^* is standard normally distributed and independent of variable U_{1k} . To verify the applicability of equation (A.98) it needs to be shown that [1] U_{2k}^* is standard normally distributed and [2] that U_{1k} and U_{2k} have a mutual correlation coefficient that is equal to ρ_{12k} . Note that this was proven earlier in the follow-up of equation (A.65). Inserting the expression for U_{2k} , given by equation (A.98), into the formula for Z_2 (equation (A.89)) gives:

$$Z_2 = \beta_2 + \alpha_{21} \left(U_{11} \rho_{121} + U_{21}^* \sqrt{(1 - \rho_{121}^2)} \right) + \dots + \alpha_{2n} \left(U_{1n} \rho_{12n} + U_{2n}^* \sqrt{(1 - \rho_{12n}^2)} \right) \quad (\text{A.99})$$

which can be written more compactly as follows:

$$Z_2 = \beta_2 + \sum_{k=1}^n \alpha_{2k} \left(U_{1k} \rho_{12k} + U_{2k}^* \sqrt{(1 - \rho_{12k}^2)} \right) \quad (\text{A.100})$$

Note that the expression for Z_1 in equation (A.89) can also be written more compactly as follows:

$$Z_1 = \beta_1 + \sum_{k=1}^n \alpha_{1k} U_{1k} \quad (\text{A.101})$$

The procedure for determining the equivalent α -values is similar to that introduced for the case of full correlation (section A.3.1.3). So again, an equivalent Z -function of the following form is derived:

$$Z^e = \beta^e + \alpha_1^e U_1 + \dots + \alpha_n^e U_n = \beta^e + \sum_{k=1}^n \alpha_k^e U_k \quad (\text{A.102})$$

The main difference with section A.3.1.3 is that the random variable U_k in equation (A.102) will represent the two variables U_{1k} and U_{2k} of the functions Z_1 and Z_2 , whereas in section A.3.1.3, U_{1k} and U_{2k} were fully correlated and therefore described by a single variable U_k . The approach in the current section is that first a separate equivalent α -value is computed for variables U_{1k} and U_{2k}^* . Subsequently, these two equivalent α -values are combined into a single equivalent α -value.

The first step is to derive the partial derivative of β^e to variable U_{1k} . Similar to section A.3.1.3 this is done by quantifying the change in β^e as a result of a small perturbation in the value of u_{1k} . β^e is related to the two Z -functions as follows:

$$\beta^e = \Phi^{-1}[1 - P\{Z_1 < 0 \cup Z_2 < 0\}] \quad (\text{A.103})$$

Substituting equations (A.100) and (A.101) into equation (A.103) gives:

$$\begin{aligned} \beta^e = & \Phi^{-1}[1 - P\{\beta_1 + \sum_{k=1}^n \alpha_{1k} U_{1k} < 0 \cup \beta_2 \\ & + \sum_{k=1}^n \alpha_{2k} \left(U_{1k} \rho_{12k} + U_{2k}^* \sqrt{1 - \rho_{12k}^2} \right) < 0\}] \end{aligned} \quad (\text{A.104})$$

A small perturbation, ε , on the mean value of u_{1k} will have the following effect on β^e :

$$\beta_{k;c}^e(\varepsilon) = \Phi^{-1}[1 - P\{Z_1 < -\alpha_{1k}\varepsilon \cup Z_2 < -\alpha_{2k}\varepsilon\rho_{12k}\}] \quad (\text{A.105})$$

This value can be computed once again with the Hohenbichler method (or alternative methods for combining two components). The equivalent α -value for variable U_{1k} can than be obtained from:

$$\alpha_k^I = \frac{\beta_{k;c}^e(\varepsilon) - \beta^e}{\varepsilon} \quad (\text{A.106})$$

Subsequently, a similar sensitivity analysis is done for U_{2k}^* . A small perturbation, ε , in the mean value of U_{2k}^* has the following effect on β_k^e (see equations (A.103) and (A.104)):

$$\beta_{k;u}^e(\varepsilon) = \Phi^{-1}[1 - P\{Z_1 < 0 \cup Z_2 < -\alpha_{2k}\varepsilon\sqrt{1 - \rho_{12k}^2}\}] \quad (\text{A.107})$$

This value can be computed once again with the Hohenbichler method (or alternative methods for combining two components). The equivalent α -value for variable U_{2k}^* can than be obtained from:

$$\alpha_k^{II} = \frac{\beta_{k;u}^e(\varepsilon) - \beta^e}{\varepsilon} \quad (\text{A.108})$$

The two derived α -values can then be combined as follows:

$$\alpha_k^e = \sqrt{(\alpha_k^I)^2 + (\alpha_k^{II})^2} \quad (\text{A.109})$$

This is the required equivalent α -value for the k^{th} random variable in the combined Z -function Z^e . Equation (A.109) can be explained as follows. The Z -functions of the two components, Z_1 and Z_2 are a function of mutually independent standard normally distributed variables $U_{11}, U_{21}^*, \dots, U_{1n}, U_{2n}^*$ (see equations (A.89) and (A.100)). For each of these variables an equivalent α -value was derived: $\alpha_1^I, \alpha_1^{II}, \dots, \alpha_n^I, \alpha_n^{II}$. This means the combined Z -function of the two components can be written as follows:

$$\begin{aligned} Z^e &= \beta^e + \alpha_1^I U_{11} + \alpha_1^{II} U_{21}^* + \dots + \alpha_n^I U_{1n} + \alpha_n^{II} U_{2n}^* \\ &= \beta^e + \sum_{k=1}^n (\alpha_k^I U_{1k} + \alpha_k^{II} U_{2k}^*) \end{aligned} \quad (\text{A.110})$$

If we compare this equation with equation (A.102) it is clear that the "new" random variable U_k replaces the pair of random variables U_{1k} and U_{2k}^* and also that the equivalent α -value, α_k^e , replaces α_k^I and α_k^{II} . This can only be done if the standard deviation of $\alpha_k^e U_k$ is equal to the standard deviation of: $\alpha_k^I U_{1k} + \alpha_k^{II} U_{2k}^*$. This is the case if we chose α_k^e according to equation (A.109).

Note: application of equation (A.109) will result in a value of α_k^e that is non-negative. For load variables this is incorrect, as they have negative α -values. Therefore, for load variables, α_k^e should be taken equal to:

$$\alpha_k^e = -\sqrt{(\alpha_k^I)^2 + (\alpha_k^{II})^2} \quad (\text{A.111})$$

If it is not known whether the k^{th} variable is a load variable, this information can be obtained by reading the sign of α_k^I and α_k^{II} .

Note: since the equivalent α -values are derived numerically, the sum of the squares of the equivalent α -values may differ from 1. In that case an additional normalization step is required:

$$\alpha_{k;final}^e = \frac{\alpha_k^e}{\sqrt{\sum_{j=1}^n (\alpha_j^e)^2}} ; k = 1 \dots n \quad (\text{A.112})$$

A.3.1.5 Systems with an arbitrary number of components

We have just considered the case of probability of failure of a parallel system of two components. In this section we extend the concept to an arbitrary number of components. Suppose we have an arbitrary system of n components. The failure probability for the system is given by:

$$P(F) = P(Z_1 < 0 \cup Z_2 < 0 \cup \dots \cup Z_n < 0) \quad (\text{A.113})$$

An example is the computation of the failure probability due to the mechanism overtopping over n defense segments. The function Z_i is the limit state function for component i and the occurrence $Z_i < 0$ indicates failure of component i .

The procedure of combining is to first combine two components, so that the problem with n components reduces to a problem with $n - 1$ components. The next step combines two components again so that the problem reduces to one with $n - 2$ components, and continues in this fashion until only one component remains, where this last component represents the entire system.

The order of the combination is important. The determination of equivalent α -values, discussed in the previous section, is an approximating method, which makes the entire combination procedure an approximating method. The accuracy of the resulting failure probability is influenced by the sequence in which the components are combined. The most accurate results are obtained by combining the most correlated components first. This is clarified by the example below with three Z -functions in Table A.4 and Figure A.23.

Table A.4: Description of the Z -functions of Figure A.23

Function	Variable		
	α_1	α_2	β
Z_1	-1	0	2
Z_2	0	-1	2
Z_3	-1	0	2.5

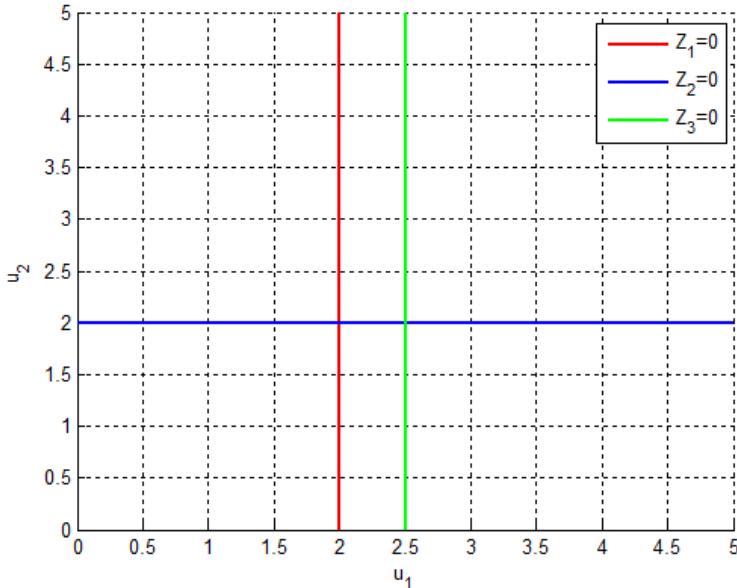


Figure A.23: Functions Z_1 , Z_2 and Z_3 of the current example

In this example, functions Z_1 and Z_3 are mutually fully correlated, whereas they are fully uncorrelated with function Z_2 . We will demonstrate that the best strategy is to first combine the two correlated Z -functions (Z_1 and Z_3). First of all the exact solution for this relatively easy example is derived. It is clear from Figure A.23 that if $Z_3 < 0 \Rightarrow Z_1 < 0$. This means:

$$P(Z_1 < 0 \cup Z_3 < 0) = P(Z_1 < 0) \quad (\text{A.114})$$

And therefore:

$$P(F) = P(Z_1 < 0 \cup Z_2 < 0 \cup Z_3 < 0) = P(Z_1 < 0 \cup Z_2 < 0) \quad (\text{A.115})$$

Since Z_1 and Z_2 are independent, the failure probability is equal to:

$$P(Z_1 < 0 \cup Z_2 < 0) = 1 - P(Z_1 \geq 0)P(Z_2 \geq 0) = 1 - \Phi(\beta_1)\Phi(\beta_2) = 1 - \Phi(2)^2 \approx 0.045 \quad (\text{A.116})$$

If we combine Z_1 and Z_3 first with the Hohenbichler method, the exact same result is obtained. If we combine Z_1 and Z_2 first, the estimated failure probability is equal to 0.0482, whereas if we combine Z_2 and Z_3 first, the probability of failure is equal to 0.0492. This demonstrates that in this example indeed the best strategy is to first combine the two components with the largest mutual correlation. To understand why this is the case, Figure A.24 shows the equivalent function, Z^e of $Z_1 \cup Z_3$. This function turns out to be exactly the same as Z_1 . Recall from equation (A.114) that this means that Z^e is an exact representation of $Z_1 \cup Z_3$.

Figure A.25 shows the equivalent function, Z^e of $Z_1 \cup Z_2$. Clearly, this function is a compromise between $Z_1 < 0 \cup Z_2 < 0$ and it is clear why this introduces some errors after combining with function $Z_3 < 0$ (for instance because Z_3 now defines part of the failure domain, whereas in the original problem statement it was redundant).

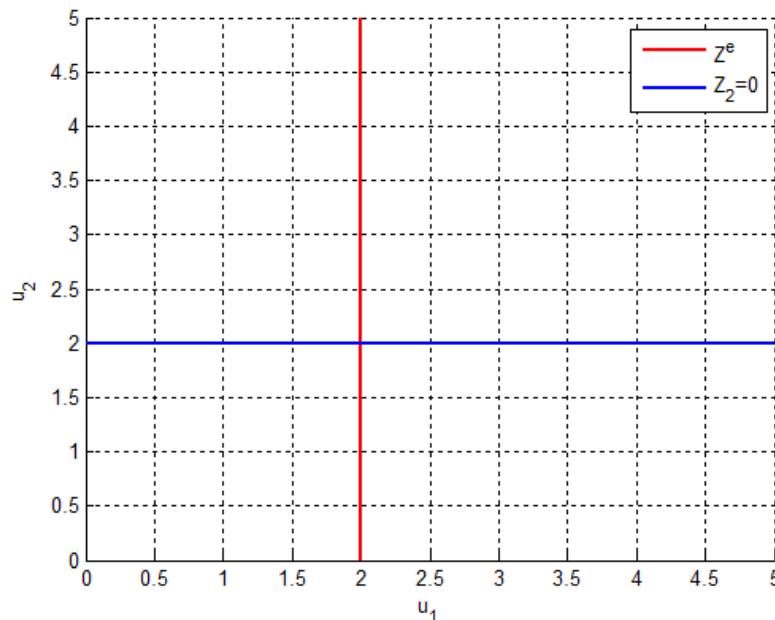


Figure A.24: Function Z_2 and the equivalent Z -function of $Z_1 \cup Z_3$.

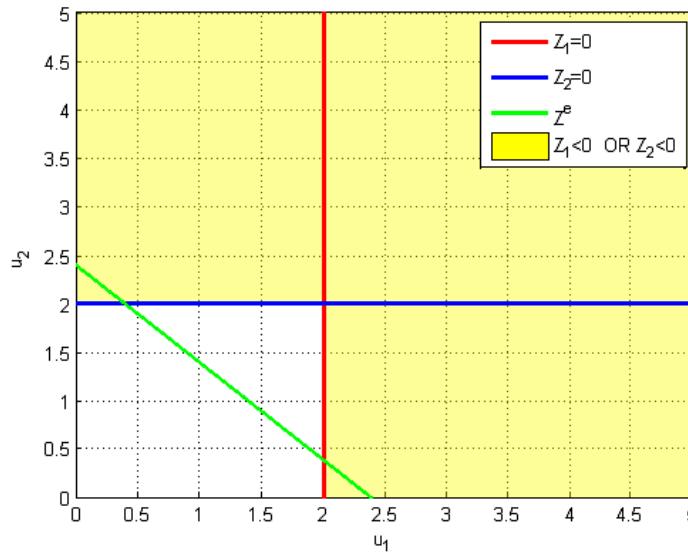


Figure A.25: Functions $Z_1=0$, $Z_2=0$ and the equivalent Z -function of $Z_1 < 0 \cup Z_2 < 0$.

Figure A.26 helps illustrate the concept of combining components with the largest mutual correlation, with an example. Shown are four components with reliability functions Z_1 , Z_2 , Z_3 , and Z_4 . Let functions Z_1 and Z_2 be the most strongly correlated. These two components are then first combined and replaced by the equivalent reliability function $Z_{1,2}^e$. For the three remaining components, the correlations between them are again computed. Consider the case where Z_3 and Z_4 are now the most correlated; the following step will be the combination of Z_3 and Z_4 , resulting in the equivalent reliability function $Z_{3,4}^e$. The final step is the combination of $Z_{1,2}^e$ and $Z_{3,4}^e$.

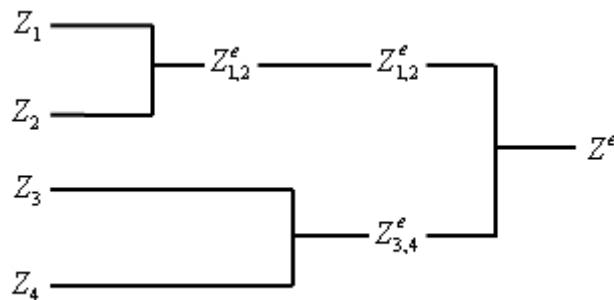


Figure A.26: Example of combining failure probabilities over four components

B Correlations between random variables

B.1 General description

In section 2.1.1 the procedure for applying statistical distribution functions of random variables was explained. In the explanation, the X -variables were assumed to be mutually independent for the sake of simplicity. In many cases, however, random variables of hydraulic load models are not mutually independent. For instance, wind speed and sea water level are correlated and the same generally can be stated for river discharges of adjacent rivers (such as the Rhine and Meuse in the Netherlands). Correlation between two random variables X_1 and X_2 needs to be taken into account in probabilistic analysis because they influence the probability of failure of the component or system under consideration.

The general approach for modeling correlation between two random variables X_1 and X_2 is to first generate correlated samples $u_{1,cor}$ and $u_{2,cor}$ from standard normally distributed variables $U_{1,cor}$ and $U_{2,cor}$.¹ The correlated samples, or realizations, are subsequently translated into realizations x_1 and x_2 of "real world" variables X_1 and X_2 through application of the procedure of section 2.1.1 (i.e. through application of the inverse CDF's of variables X_1 and X_2). The procedure is depicted in the figure below. The horizontal part of this figure is the exact same procedure as depicted in Figure 2.2. The correlation model can therefore be considered as a pre-processing of the procedure in which the distribution functions are applied.

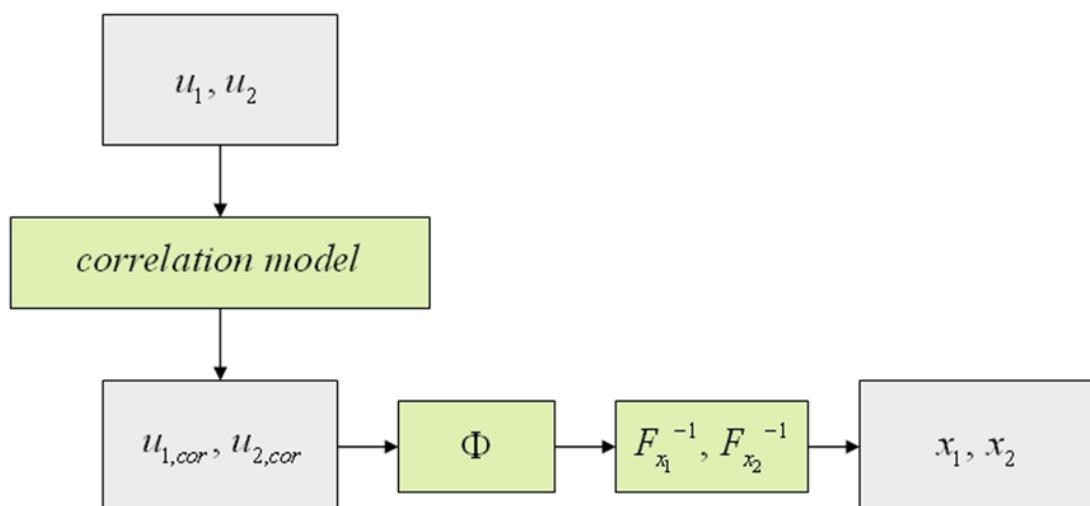


Figure B.1: Procedure for determining a load variable associated with randomly selected standard normally distributed variables for the case of correlated variables.

Since $U_{1,cor}$ and $U_{2,cor}$ are correlated variables, X_1 and X_2 are also correlated. Furthermore, since $U_{1,cor}$ and $U_{2,cor}$ are standard normally distributed and the translation from $U_{1,cor}$ and $U_{2,cor}$ to X_1 and X_2 is done in the exact same way as described in section 2.1.1, it is automatically taken care of that X_1 and X_2 are distributed according to their prescribed distribution functions F_{X_1} and F_{X_2} . The remainder of this section therefore focuses on the first part of the procedure: the generation of samples $u_{1,cor}$ and $u_{2,cor}$ of correlated variables $U_{1,cor}$ and $U_{2,cor}$.

¹Note that u_1 and u_2 are strictly speaking only "realizations" if a Monte Carlo procedure is applied, see section A.2. For methods like FORM and numerical integration, u_1 and u_2 are strategically selected values, not samples from a simulation of a distribution function. However, this fundamental difference in interpretation of u_1 and u_2 has no influence on the applied methods as described in the current section.

The generation of $U_{1,cor}$ and $U_{2,cor}$ starts with the generation of realizations u_1 and u_2 of independent standard normally distributed random variables U_1 and U_2 . Subsequently, u_1 is transformed into a sample v of variable V with distribution function $F_V(v)$. The transformation is done in similar style as explained in section 2.1.1, i.e. by making sure the probability of (non-)exceedance of u_1 and v are equal:

$$\Phi(u_1) = F_V(v) \Rightarrow v = F_V^{-1}(\Phi(u_1)) \quad (\text{B.1})$$

The Φ in this equation is the standard normal cumulative distribution function and $F_V(v)$ can be any cumulative distribution function, such as the normal, uniform or exponential distribution distribution. Subsequently, a sample w is introduced that is dependent on v and on a sample u_2 from a second standard normally distributed variable U_2 :

$$w = G(v, u_2) \quad (\text{B.2})$$

The G in this equation is a function that determines the correlation between v and w . Subsequently, v and w are transformed back into samples $u_{1,cor}$ and $u_{2,cor}$ of standard normally distributed variables $U_{1,cor}$ and $U_{2,cor}$ by using the inverse CDFs or probabilities of (non-)exceedance. This leads to a $u_{1,cor}$ that identical to u_1 , but $u_{2,cor}$ will be different from u_2 because of the use of the correlation function G .

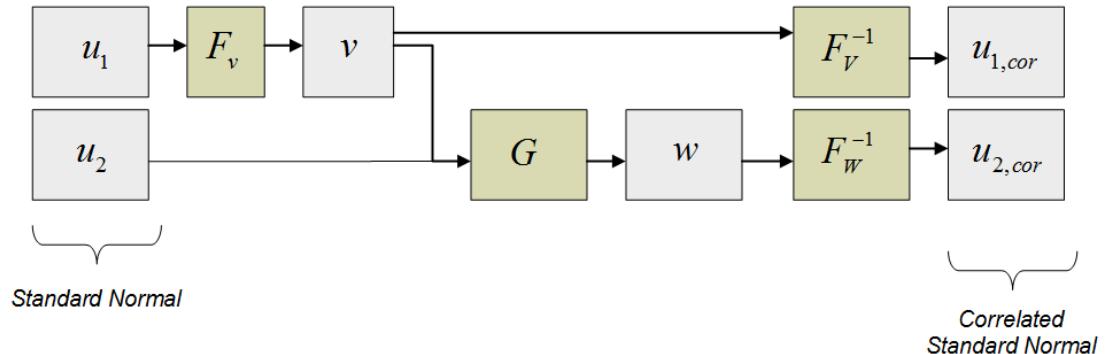


Figure B.2: Procedure for samples $u_{1,cor}$ and $u_{2,cor}$ of correlated standard uniform random variables $U_{1,cor}$ and $U_{2,cor}$.

The function G is essentially the correlation model for variables X_1 and X_2 . Many different functions G can be used (note: G is not a distribution function), also because the subsequent transformations F_V^{-1} guarantee that X_1 and X_2 are distributed according to the pre-defined distribution functions F_{X_1} and F_{X_2} . Naturally, function G should be chosen such that it reproduces the observed correlation between variables X_1 and X_2 as well as possible. For a more detailed background on the derivation and application of correlation models in flood risk analysis, the interested reader is referred to the paper of Diermanse and Geerse (2012).

B.2 Correlation models

In this section the bi-variate Gaussian correlation model and its generalization, the Gaussian copula model, are described.

The bi-variate Gaussian correlation model has been implemented for practical purposes and it is described in section B.2.1. The correlation model is a special case of the Gaussian copula model: the Gaussian copula model with 2 variables gives the same results as the bi-variate Gaussian correlation model. The Gaussian copula model for ≥ 2 random variables is described in section B.2.2.

B.2.1 Gaussian correlation model

This section describes the bi-variate Gaussian correlation model that can be used to describe a simple correlation between standard normal random variables. Given is the following correlation matrix C with correlation coefficient $\rho \in [-1, 1]$:

$$C = \begin{bmatrix} 1.0 & \rho \\ \rho & 1.0 \end{bmatrix} \quad (\text{B.3})$$

The 'lower' Cholesky decomposition of this matrix is (i.e. $P \times P^T = C$):

$$P = \begin{bmatrix} 1.0 & 0.0 \\ \rho & \sqrt{1 - \rho^2} \end{bmatrix} \quad (\text{B.4})$$

Given an uncorrelated sample of standard normal random variables u_1 and u_2 , the correlated sample $u_{1,cor}$ and $u_{2,cor}$ is derived as follows:

$$[u_{1,cor}, u_{2,cor}] = [u_1, u_2] \times P^T = [u_1, u_2] \times \begin{bmatrix} 1.0 & \rho \\ 0.0 & \sqrt{1 - \rho^2} \end{bmatrix} \quad (\text{B.5})$$

This simplifies to:

$$u_{1,cor} = u_1 \quad (\text{B.6})$$

and

$$u_{2,cor} = u_1\rho + u_2\sqrt{1 - \rho^2} \quad (\text{B.7})$$

The following figure pictures the uncorrelated and correlated samples for $\rho = 0.5$.

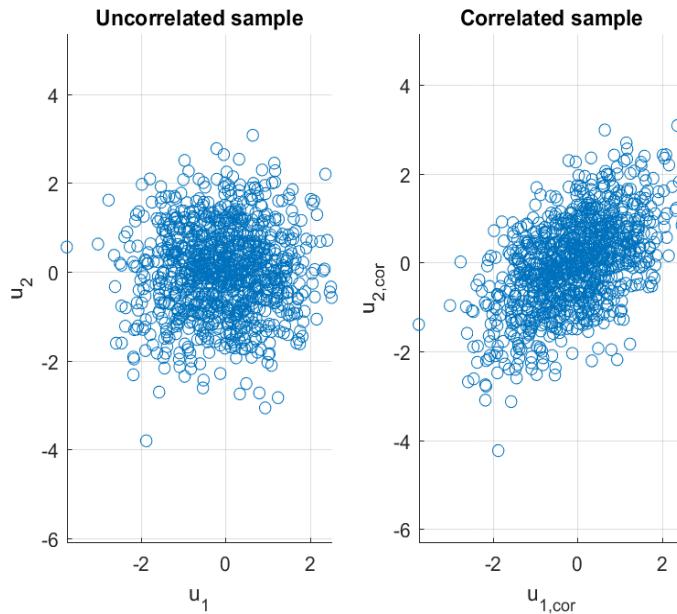


Figure B.3: Uncorrelated (left) and correlated (right) samples according to the Gaussian correlation model assuming $\rho = 0.5$.

Furthermore, the following holds:

- ◊ When $\rho = 0.0$ then the correlated sample is equal to the uncorrelated sample.
- ◊ When $\rho = 1.0$ then $u_{2,cor} = u_1$ (complete correlation).
- ◊ When $\rho = -1.0$ then $u_{2,cor} = -u_1$ (reverse complete correlation).

B.2.2 Gaussian copula model

This section presents a Gaussian copula model, which can be used to describe correlation between ≥ 2 variables in the U -space. The correlation coefficients are in this case stored in a correlation matrix.

A correlation matrix is a table showing correlation coefficients (i.e. product moment correlations) between variables. The diagonal of the table is always a set of ones, because the correlation between a variable and itself is always 1.0. The matrix is also symmetric – the correlation coefficient between variables X and Y is equal to the correlation coefficient between variables Y and X . Moreover, the matrix must be positive definite, that is:

$$x^T \cdot C \cdot x > 0 \text{ for all } x \in \Re^n \setminus \{0\} \quad (\text{B.8})$$

where C is a correlation matrix with size $n \times n$ and T denotes the transpose.

In the Probabilistic Library, it is possible to describe correlations between random variables in a correlation matrix. The procedure to generate correlated samples u_{cor} is as follows (see Diermanse *et al.* (2014)):

- 1 Derive a matrix P for which $P \cdot P^T = C$ through Cholesky decomposition of correlation matrix C with size $n \times n$ (see Strang (1982)).
- 2 Sample values u_1, \dots, u_n from the standard normal distribution function and store the values in a $1 \times n$ vector u .
- 3 The sample of the correlated standard normally distributed random variables u_{cor} is then defined as follows: $u_{cor} = u \cdot P^T$.

In the Probabilistic Library a special care is taken of correlation matrices with off-diagonal ± 1.0 elements. Such matrices usually do not satisfy the positive definite requirement. In order to perform calculations with such matrices, a two-step procedure is applied, during which the original matrix is decomposed into two matrices: (1) a matrix without the off-diagonal ± 1.0 elements and (2) a matrix with 0.0 and ± 1.0 entries that indicate the positions of the off-diagonal ± 1.0 elements in the original matrix. The Cholesky decomposition is performed on the first matrix, highly increasing the chance that the positive definite requirement will be satisfied. The vector u is first multiplied with the Cholesky decomposition matrix entailing a correlated sample without the full correlation cases. The resulting sample vector is then multiplied by the second matrix assuring that the same values are assigned to the fully correlated variables. The procedure is explained in more detail in the following example.

Consider the following correlation matrix C with $\rho_{1,2} = \rho_{2,1} = 1.0$:

$$C = \begin{bmatrix} 1.0 & \rho_{1,2} & \dots & \rho_{1,j} & \dots & \rho_{1,n} \\ \rho_{2,1} & 1.0 & \dots & \rho_{2,j} & \dots & \rho_{2,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{i,1} & \rho_{i,2} & \dots & 1.0 & \dots & \rho_{i,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{n,1} & \rho_{n,2} & \dots & \rho_{n,j} & \dots & 1.0 \end{bmatrix} \quad (\text{B.9})$$

Matrix C is decomposed into matrix C_1 and C_2 . Matrix $C = C_1$ except for off-diagonal values found in the second row and column. These values are all equal to 0.0. Matrix C_2 is an identity matrix except for the second row and column where $\rho_{2,1} = 1.0$, $\rho_{i,2} = 0.0$ for $i = 1\dots n$ and $\rho_{2,j} = 0.0$ for $j = 2\dots n$.

$$C_1 = \begin{bmatrix} 1.0 & \mathbf{0.0} & \dots & \rho_{1,j} & \dots & \rho_{1,n} \\ \mathbf{0.0} & 1.0 & \dots & \mathbf{0.0} & \dots & \mathbf{0.0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{i,1} & \mathbf{0.0} & \dots & 1.0 & \dots & \rho_{i,n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{n,1} & \mathbf{0.0} & \dots & \rho_{n,j} & \dots & 1.0 \end{bmatrix} \quad (\text{B.10})$$

$$C_2 = \begin{bmatrix} 1.0 & \mathbf{0.0} & \dots & 0.0 & \dots & 0.0 \\ \mathbf{1.0} & \mathbf{0.0} & \dots & \mathbf{0.0} & \dots & \mathbf{0.0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.0 & \mathbf{0.0} & \dots & 1.0 & \dots & 0.0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.0 & \mathbf{0.0} & \dots & 0.0 & \dots & 1.0 \end{bmatrix} \quad (\text{B.11})$$

Assuming that P is the Cholesky decomposition of matrix C_1 , then the correlated sample u_{cor} is derived as follows:

$$u_{cor} = u \cdot P^T \cdot C_2^T \quad (\text{B.12})$$

C Conversions

C.1 Background

C.1.1 Floating point computations

In the Probabilistic Library three types of real (floating point numbers) are defined:

- ◊ `integer, parameter :: sp = selected_real_kind(6, 37) !< Single precision`
- ◊ `integer, parameter :: dp = selected_real_kind(15, 307) !< Double precision`
- ◊ `integer, parameter :: qd = selected_real_kind(33, 4931) !< Quadruple precision`

The quadruple precision (`kind=qd`) is only applied for the testing. Types `sp` and `dp` correspond to the IEE 754 standard for single and double precision:

- ◊ single precision: 32-bit number with 1 sign bit for positive/negative, 8 bit exponent (in excess-127) and significant precision of 23 bits. This gives a range of $\pm 1.18 \cdot 10^{-38}$ to $\pm 3.4 \cdot 10^{38}$.
- ◊ double precision: 64-bit number with 1 sign bit for positive/negative, 11 bit exponent (in excess-1023) and significant precision of 52 bits. This gives a range of $\pm 2.23 \cdot 10^{-308}$ to $\pm 1.80 \cdot 10^{308}$.

Note: the Probabilistic Library usually deals with very small probabilities. Therefore all computations are performed in the double precision.



Small rounding errors are always made when calculating with floating point numbers. Floating point numbers consist of a finite number of bits and so not all real numbers (infinitely many) can have their own unique floating point representation. There are numbers that differ so little from each other that they are equal in floating point notation. The term **machine precision** is a measure of the maximum error made in this rounding and is often defined as the smallest number ϵ for which it still holds that $1 + \epsilon > 1$.

- ◊ machine precision for single precision reals $\approx 1.19e^{-07}$
- ◊ machine precision for double precision reals $\approx 2.22e^{-16}$

In practice, the first 7 decimal places are significant for single precision numbers, and 16 for double precision. Because of this finite accuracy, calculations that are identical mathematically cannot produce the same calculation result. Sometimes the order in which a calculation is performed can make a big difference in the final answer. Another possibility for the occurrence of (large) differences is to subtract two floating point numbers that are almost equal to each other. The significant figures are lost in the processing and the result contains less or possibly no significant decimal anymore.

C.1.2 The (inverse) cumulative distribution function (CDF)

The cumulative distribution function (CDF) or a random variable X is a function defined as follows, for $x \in \mathbb{R}$:

$$F_X(x) = P(X \leq x) \tag{C.1}$$

In case of continuous random variables, the CDF can be written in terms of the probability density function f_X :

$$F_X(x) = \int_{-\infty}^{\infty} f_X(t) dt \quad (\text{C.2})$$

The inverse function of the CDF is by definition the function that returns the value of x for which the probability is less than or equal to a given probability. The inverse CDF is also called **percent-point** function or **quantile** function. In terms of F_X , the inverse CDF returns x for the given p so that:

$$F_X(x) := P(X \leq x) = p \quad (\text{C.3})$$

C.1.3 The (inverse) of the standard normal distribution function

The CDF of the standard normal distribution function, often denoted by Φ , is defined as follows:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \quad (\text{C.4})$$

There is a direct relation between Φ and the error function ($\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$):

$$\Phi(x) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (\text{C.5})$$

$$\text{erf}(x) = 2\Phi(x\sqrt{2}) - 1 \quad (\text{C.6})$$

The inverse function of the CDF of the standard normal distribution is called the **probit** function. By definition, the following must hold:

$$\Phi(\text{probit}(p)) = p \quad (\text{C.7})$$

$$\text{probit}(\Phi(x)) = x \quad (\text{C.8})$$

Using equation C.6, the probit function can be written in terms of the inverse of the error function $\text{erf}^{-1}(x)$:

$$\text{probit}(p) = \sqrt{2} \cdot \text{erf}^{-1}(2p - 1) \quad (\text{C.9})$$

There are no closed expressions for the error function and the probit function. Because these functions are often used in (statistical) calculations, there are several numerical approaches available.

C.2 Conversion of reliability index to probability (and vice versa)

Computation techniques applied in the system reliability result in reliability index β (section 5.2). For real life applications, it is often required to translate the reliability index to the (non)failure probability and vice versa.

To calculate failure probability q from reliability index β the following equation is applied:

$$q = \Phi(-\beta) \quad (\text{C.10})$$

Non-failure probability p on the other hand is derived as follows:

$$p = \Phi(\beta) \quad (\text{C.11})$$

The following equation yields the reliability index β given failure probability q :

$$\beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^q e^{-\frac{1}{2}t^2} dt \quad (\text{C.12})$$

Note: the above conversions require a numerical approximation of the probit function. In the Probabilistic Library this is done using the methods of [Wichura \(1988\)](#) and [Hart et al. \(1968\)](#). Both methods are double precision.



References

- Au, J. L., S.K.; Beck, 2001. "Estimation of small failure probabilities in high dimensions by subset simulation." *Probabilistic Engineering Mechanics* 16 (4): 263-277.
- Bjerager, P., 1988. "On computation methods for structural reliability analysis." In: *Frangopol DM, editor. New directions, in structural system reliability* pages 52-67.
- Deltas, 2023. *Technical Reference Manual of Hydra-Ring 23.1.1.* Tech. rep., Deltas, 11209269-008-GEO-0004.
- Diermanse, F., D. Carroll, J. Beckers, R. Ayre and J. Schuurmans, 2014. "A Monte Carlo framework for the Brisbane river catchment flood study." In *Hydrology and Water Resources Symposium 2014*, pages 62-69. Barton, ACT: Engineers Australia.
- Diermanse, F. and C. Geerse, 2012. "Correlation models in flood risk analysis." *Reliability engineering and system safety* 105: 64-72.
- Forbes, 2010. *Statistical Distributions.* John Wiley and Sons.
- García, O., 1981. "Simplified method-of-moments estimation for the Weibull distribution." *New Zealand Journal of Forestry Science* 11: 304-306.
- Grimmett, G. and D. Stirzaker, 1983. *Probability and random processes.* Oxford science publications.
- Hart, J., W. Cheney, C. Lawson, H. Maehly, C. Mesztenyi, J. Rice, H. Thacher and C. Witzgall, 1968. *Computer Approximations.* Wiley.
- (<https://stats.stackexchange.com/users/2392/probabilityislogic>) probabilityislogic, 2013. "Estimating parameters of Student's t-distribution." Cross Validated. URL:<https://stats.stackexchange.com/q/63861> (version: 2013-07-10).
- Jongejan, 2012. *Het lengte-effect in een statistisch homogeen vak.* Tech. rep., Jongejan RMC.
- Lynch, S., 2007. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* Springer.
- Melchers, R., 2002. *Structural reliability analysis and prediction.* John Wiley and sons.
- Papaioannou, I., W. Betz, K. Zwirglmaier and D. Straub, 2015. "MCMC algorithms for Subset Simulation." *Probabilistic Engineering Mechanics* 41: 89-103. DOI: <https://doi.org/10.1016/j.probengmech.2015.06.006>, ISSN 0266-8920, URL <https://www.sciencedirect.com/science/article/pii/S0266892015300205>.
- Powell, M., 1994. "A direct search optimization method that models the objective and constraint functions by linear interpolation." *Advances in Optimization and Numerical Analysis, eds. S. Gomez and J.-P. Hennart* pages 51-67.
- Rosenblatt, M., 1952. "Remarks on a multivariate transformation." *Ann. Matth Stat.* 23: 470-472.
- Rosenblatt, M., 2011. "An adaptive directional importance sampling method for structural reliability." *Probabilistic Engineering Mechanics* 26: 134-141.
- Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, 2010. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Computer Physics Communications* 181 (2): 259-270. DOI: <https://doi.org/10.1016/j.cpc.2009.09.018>, ISSN 0010-4655, URL <https://www.sciencedirect.com/science/article/pii/S0010465509003087>.

- Strang, G., 1982. *Linear algebra and its applications*. Harcourt, Brace, Jovanovich.
- Waarts, P., 2000. *Structural reliability using finite element analysis; an appraisal of DARS: Directional Adaptive Response Surface Sampling*. Delft University of Technology.
- Wichura, M., 1988. "Algorithm AS 241: The Percentage Points of the Normal Distribution." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* pages 477-484.